EPM/RT-85-1

A TAXONOMIC ANALYSIS OF THE FURNITURE MANUFACTURING INDUSTRY

Laurent Villeneuve, Eng., M.Eng., professor
D.-Bongs Lainjo, B.Sc., M.Sc.A.

Département de Génie Industriel

Ecole Polytechnique de Montréal

JANUARY 1985

REMERCIEMENTS

Ce projet a été réalisé grâce au soutien financier du Conseil de recherches en sciences et en génie du Canada (A-8803).

Dépôt légal, 1er trimestre 1985
Bibliothèque nationale du Québec
Bibliothèque nationale du Canada

Pour se procurer une copie de ce document, s'adresser au:

Service de l'édition
Ecole Polytechnique de Montréal
Case postale 6079, Succursale A
Montréal (Québec) H3C 3A7
(514) 340-4903

Compter 0,05$ par page (arrondir au dollar le plus près), plus 1,50$ (Canada) ou 2,50$ (étranger) pour la couverture, les frais de poste et la manutention. Régler en dollars canadiens par chèque ou mandat-poste au nom de l'Ecole Polytechnique de Montréal. Nous n'honorerons que les commandes accompagnées d'un paiement, sauf s'il y a eu entente préalable, dans le cas d'établissements d'enseignement ou d'organismes canadiens.

Abstract

The current article is a product of an exploratory study of furniture manufacturing industry in Quebec. Numerical taxonomic techniques are applied in an effort to (1) identify the parameters that can be suitably used to describe a set of firms; (2) quantify and classify these parameters in order of importance and (3) factor analytically situate or classify the variables which make up this industry. The companies are then classified. The classification is based on activities or variables which collectively define the sectors' production system. A field study, based on interviews (questionnaires), is then presented and analysed in an attempt to develop or identify a functional production system applicable to a given industry.

Résumé

Cet article est le résultat d'une enquête sur l'industrie du meuble au Québec. Des techniques de taxonomie ont été appliquées pour (1) identifier les paramètres qui peuvent décrire un groupe d'entreprises; (2) quantifier et classifier ces paramètres par ordre de priorité et (3) situer analytiquement les facteurs ou classifier les variables qui constituent cette industrie. Les entreprises ont été ainsi classifiées. Cette classification est basée sur les activités ou variables qui globalement définissent les systèmes de production de ce secteur. Les résultats de l'étude, basés sur des entrevues par questionnaire sont présentés et analysés dans le but de développer ou d'identifier un système de production fonctionnel applicable à cette industrie.

Table of contents

	Page
Remerciements	i
Abstract	ii
Résumé	iii
Table of contents	iv
Introduction	1
Sample and survey method	5
Variable definition	6
Quality control	6
Limitations	7
Variable correlation	9
Correlation matrix	10
Analysis of variables	11
Multicollinearity	12
Cluster analysis	13
Results	15
Selected variables	16
Literature	17
Theoretical background	18
Factor extraction of principal components	20
Transformed dimensions	26
Observations	30
Homogeneous firms	32

Original dimensions 33

Apparent attributes 34

Latent attributes 35

Conclusions 37

Appendix 1 - Cluster representation of variables 38

References 45

Introduction

Many industrial experts will agree with the statement that the last couple of years have been of great concern to managers and entrepreneurs. This is confirmed by the tremendous, diversified and complex nature of the problems they have continuously had to solve in their respective industries.

An objective look at a production system indicates a spectrum of coordinated or uncoordinated set of operations discrete or continuous generating some form of output or final product. The parameters which define this system are both quantitative and qualitative in structure. These vary from basic raw materials to manpower and complex machinery. The qualitative and quantitative characteristics of such a system generate two major problems of concern-academic and practical. An even more delicate aspect is the task both engineers and managers alike have to accomplish in our contemporary industries.

The degree of complexity and heterogeneity of operations in the different industrial sectors tends to render analytical solutions-deterministic and probabilistic (with complementary judgements) invaluable. The advent of computers, however has both minimized (for the receptive managers) and magnified (for the phobic managers) problems in some organizations. The problem minimization is confirmed by the realization of this research project which involved the analysis of copious variables (more than 600 variables initially).

What generated this article is one of the major problems in the study of industrial sectors nowadays. That is; how to efficiently and reliably utilize tons of data available in their decision making processes. Should every bit of information be considered vital? If not what should be used and what should be declared redundant? One can almost certainly say that many major analytical problems require the categorization of information or data. In forecasting for example, the importance attributed to variables cannot be over emphasized. A careful identification of both dependent and independent variables leads to a more reliable forecast. The question then arises, how can the variables used be scientifically classified and justified? What are the risks of operating with either too much or too little information. Sokal (1966) presents it as follows:

> "Classification is one of the fundamental concerns of science. Facts and objects must be arranged in an orderly fashion before their unifying principles can be discovered and used as a basis for prediction. Many phenomena occur in such variety and profusion that unless some system is created among them, they would be unlikely to provide any useful information".

With the arrival of a new technology-numerical taxonomy, the researcher is more equiped to develop the requisite and necessary conceptual framework for data classification in this information oriented era.

A look at the current research trends indicates that many different specialized groups (as confirmed in different professional journals) have for a while concentrated on proposing solutions to some of these problems.

Scientifits , economists, psychologists etc. have built and tested models in an attempt to provide solutions to some of the perennial problems in industry. For example, models on decision matrices and decision trees proposed by Brown (1970) currently used by some managers. In some of his works, he grouped items into families and established a coordinated policy for a determistic demand.

This article is aimed at generating some input towards the solution of some of the above mentionned problems. The study can be broken down into four stages:

(1) Original data matrix: this represents the initial data collected through interviews with companies.

(2) Coefficient matrix: aimed at finding a correlation (interrelationship) between variables.

(3) Cluster analysis: examines the similarity in the group of variables analysed in (2). The variables are then grouped to form clusters (families). From each cluster, one representative variable is chosen.

(4) Factor analysis: uses the collection of variables chosen in (3) in order to determine new categories or factors. This analysis is useful in the following ways:

(1) It can be used to establish relationships among variables as indicated by latent factors.

(2) It identifies obvious relationships that may exist among observed data.

(3) A good data reduction technique when information needs grouping.

(4) It empirically groups variables into homogeneous entities. These will be explained further in this report.

In the next section, the survey methodology, quality control, limitations and literature are presented. This is followed by a practical illustration of cluster analysis. The subsequent section factor analysis (the nucleus of the research) the variables generated by the cluster analysis. The forth section attempts to classify the companies into homogeneous groups. The article ends with some concluding observations.

Sample and Survey method

The choice of the sample size for this research was influenced predominantly by the needs of the Quebec Ministry of Industry and Commerce. Hence the sample frame under study is a group of household wood furniture manufacturing firms in the province of Quebec. No conventional random methodology was used in determining the required size. As a result, the question of whether the sample is representative or not can only be answered based on the fact that the firms surveyed made up for more than 90% of the universe if limited to Quebec AFMQ (Association des Fabricants de Meubles du Québec) only. Initially, the questionnaire was applied to 77 firms. The sectors are made up of the following types of furnitures: household wood furniture, lamp, office furniture, upholstered and others. Forty-one (41) of these companies belong to the household wood furniture manufacturing subgroup - sample size for this research.

The strong existence of homogeneity was also a prime factor in the choice. Finally careful consideration was given to the number of missing values present in each sector (unanswered questions). This was checked with a calculation with Nie, Hull, Jenkins, Steinbrenner, Bent (1975) Pearson Product Moment Correlation coefficients which indicated 41 firms for all pairs of variables.

Variable definition

For convenience and easy reading, it was felt necessary to give short and meaningful names-variables - to the number of activities or operations the industries were involved in. For example the variable "PPHY010" identified the question referring to the set up frequency with respect to machinery. The variable "PORG016" identified operations related to the amount of control applied during manufacturing. For example which are the automatic and manual operations involved? The possibilities were automatic control (defection) and similar intervention; automatic detection and human intervention; human control and intervention. A complete list of all the operations is presented in appendix 1.

Quality control

In most data analytic researches, the influence of error is a prime concerned to those trying to study data statistically. This feeling is justified especially when one considers the different transformation processes and individuals involved during these operations. A strong existence of errors in data can contribute enormously in distorting the findings of a research thereby yielding wrong results. It seems as if the problems created by errors will still be around for quite a while since very little work exists in literature (Rummel (1972) and Naroll (1962)) aimed at minimizing them.

In this project like many others, the sources of errors ranged from clerical to methodological. The clerical errors - coding and keypunching - were checked before coding was performed. The coded sheets were verified

after each successful coding. This was followed by keypunching the data on cards which were again verified before any programming was done. A final check on the data was performed. Here all the variables and their coded values were printed for statistical verification; (see Nie, Hull, Jenkins, Steinbrenner, Bent (1975)). This output also served as a check to the program reliability. In addition, an observation of the data distribution on each variable was made for reasonableness. These quality control procedures were very instrumental in minimizing many errors. The methodological errors arising during the interviews could not easily be controlled. The only major control was the use of two interviewers who were versed with the operations in the firms under study.

Limitations

If the requirements for performing factor analysis had been rigidly followed, this project would never have been realised. One major consolation was the fact that many research projects with similar problems have been performed (Harman (1976), Comrey (1973) and Rummel (1972)) with success. In this study, the factor analysis input is a 41 by 39 matrix. This is already contrary to the requirement (see Lawlis, Chatfield (1974)) of using a sample size twice the variable size. There are two major reasons to this. The first was economical. For financial reasons, it was felt that having 41 firms was the optimum feasible. The second reason was scarcity. That is if the money were available, there were not as many wood furniture manufacturing firms in the AFMQ Province of Quebec to satisfy the sample requirements.

The second major problem was that of measurement scales. The ideal scale for factor analysis is metric (Sheth (1971)). Although this played an important role in the choice of variables, it was still not possible to have all variables on metric scales. Hence there exists a conglomeration of metric and non-metric measurement scales.

There was yet another problem arising from variable distribution. Factor analyses require all its variables to be normally distributed. In this study, normality was only assumed (see Meyer (1975)). Hence no transformations were performed.

The type of data slice applied is an R-Factor Analysis i.e. data matrix consisting of variables (columns) versus firms. The model and technique used are component factor analysis and oblique rotation. The former because it concentrates on total variance (common + unique + error) and the latter for its absence of orthogonality among factors (Green (1978)) naturally more realistic. It must be pointed out however, that the ideal approach (Comrey (1973)) of designing a questionnaire based on a predetermined model was not followed. Here, factor analysis was applied as a by-product which nevertheless produced interesting results. That is the presence of data prompted an exploratory statistical analysis.

Variable correlation

The purpose of this section is to examine and find out the possible relationships that exist among the variables. Since the presence or absence of linearity could not be established directly, it was felt that an observation of the correlation matrix of pairs of variables would be satisfactory. This calculation of correlations served two purposes. The first as an indication of linearity which is one of the basic requirements of factor analysis. The second was to use this relationship in establishing similar groups of variables thus reducing or completely eliminating any existing types of redundancies. The clusters of variables identified in this process are then re-examined with an end result of selecting the suitable and appropriate variables to be factor analysed.

Correlation matrix

Since the main criterion which prompted the use of factor analysis was the relations between the variables, a close look at the preliminary operations involved was deemed necessary. The matrix of 99 variables analysed provides measures of statistical association - the degree to which two or more variables tend to vary simultaneously. The correlation coefficients - elements of the correlation matrix - were calculated by the Pearson product-moment method.

Analysis of variables

An observation of the correlation matrix readily indicates that some variables have very few, if at all, correlation coefficients with other variables. An example like "INTR001" (supplier dependant related variable) has only four correlation coefficients with the other 98 variables greater than an absolute value 0.30. Variables with smaller correlations stand as almost distinct attributes of industries. On the other hand, variables with opposite - larger - correlation values represent central attributes of firms by abstracting a wide range of variation among companies in their attributes. An analysis of some of the problems generated by high variable correlations in factor analysis is presented in the next section.

Multicollinearity

The existence of high correlations among sets of independent variables - multicollinearity - can be of great concern in many data analysis operations - factor, regression analysis etc. This draw back created two major problems in this study.

(1) The presence of redundancy. That is the selection of two or more variables which have the same attributes mathematically - perform the same function.

(2) Indeterminacy of factor solution. That is the impossibility of calculating the determinant used in computing the inverse of the correlation matrix. This is caused by the existence of one or more combinations of variable equations being a linear combination of others arising from (1) above.

The two possible solutions to these problems were:

(a) Ignore the presence of multicollinearity;

(b) Delete one or more of the "offending" variables.

The first solution was not considered because that meant terminating the research. The second, however, was not as obvious as it seems. The solution which is explained in the next section furnished an output of clusters of variables. Empirically, if a group of variables - clusters - is strongly collinear, a member of the group whose measurement reliability and theoretical importance is higher was selected.

Cluster analysis

Two cluster analysis methods - connectiveness and diameter - were applied. The algorithm used was based on that proposed by Johnson (1967). With the similarity function - correlation coefficient - on 99 variables as input, the hierarchical clustering algorithm is as follows:

1. Clustering Co, with value 0 yields the weak clustering
2. With the assumption of the Clustering C_{j-1} given with d - similarity function - defined for all objects or clusters in C_{j-1}. If α_i is equal to minimum non zero matrix entry, merge the pair of objects and/or clusters with distance α_j to generate C_{j1} of value α_j.
3. A new similarity function for C_j is generated as follows: if x and y are clustered in C_j and not in C_{j-1} (i.e. $d(x,y) = \alpha_j$), we define the distance from the cluster (x,y) to any third object or cluster, X, by $d((x,y),Z) = \text{MIN}\ (d(x,z),\ d(y,z))$ (1) for connectiveness method

   and

   $d((x,y),Z) = \text{MAX}\ (d(x,z),\ d(y,z))$ (2) for diameter method.

In expression (1) if x and y are objects and/or clusters in C_{j-1} not clustered in C_j, d(x,y) stays unchanged. A similarity function d for G was obtained.

Expression (2) is when x and y are two objects and for clusters of C_{j-1} which cluster in C_j, and Z is any third object of cluster of C_j.

4. Steps 2 and 3 are repeated until the strong clustering is finally obtained.

The version of the program by U.P.P.C. whose input was absolute values of the correlation matrix, used these variables for computation (Johnson (1967)).

Results

Even though these analyses are restricted to the diameter method, both methods were applied. The diameter method was used because its groupings based on maximum distance computations between groups or clusters of variables identified meaningful entities.

An evaluation of the different variable clusters was performed in order to find out which types of variables grouped together. In some cases similar questions were put differently and identified as different variables. For example "EXTR003" and "EXTR009" (see Appendix 1) are two different variables refering to the same description. These, like some others were identified under the same cluster. Another outstanding cluster was cluster number 22 where INTR005 and INTR011 (Appendix 1) where similarly identified as belonging to the same group. One cluster that looked very unreal was cluster number 34 where there was perfect correlation coefficient of 1 between PPHY015 and RHUM001 (see Appendix 1). Since these correlation coefficients do not indicate causality, all that could be concluded was that generally correlated variables are variables which tend to vary in the same direction.

Selected variables

After a series of judgemental expertise and copious variable combinations, 38 variables identified in appendix 1 with "*" were selected. These variables were considered more meaningful in terms of function and scales. Once a variable was recommended, the second consideration was its scale - metric or non metric. During this process, metric scaled variables were normally given more priority. As a result, almost all metric scaled variables among the 99 original variables were chosen and subsequently used as the input variables in factor analyzes. Having identified groups which depict basic attributes of the original correlation matrix, one can proceed to factor analyze the 38 variables as described in the next section.

Literature

Factor analysis has for quite some time received increased attention. This is confirmed by the existence of copious reliable computational procedures Harman (1976), Comrey (1973), Rummel (1970) and Kendall (1968) and the proposals of many proofs and extensions. While there exists so much to offer in literature, it is interesting to note that the emperical applications have been extensively tested in only some fields (Harman (1976), Rhodes (1937),Godfrey, Fiedler, Hall (1958), Krumbein, (1937), Imbrie (1963)). This was a very encouraging observation. However all efforts to identify an application similar to this (for comparison purposes) proved abortive. As a result, this could be regarded as a precedent geared at generating more enthusiasm in the field of production research. The computational hardships encountered in the early stages have been enormously reduced by the wide spread availability of computers and related software packages. Some of those used were by Nie, Hull, Jenkins,Steinbrenner (1975), Dixon (1975) and Howard, Harris (1966). In this section, an attempt is made at explaining the algebraic model followed by an example aimed at illustrating the theory as presented by Comrey (1973).

THEORETICAL BACKGROUND

In many data analysis procedures, the input is generally the raw data as obtained in the field. Factor analysis, however, remains one of the few exceptions. Here, the input is a matrix of correlations among data variables which ultimately produces a matrix of factor loadings. These loadings can generally be regarded as correlations between data variables and artificial constructs called "factors" or independent dimensions in the orthogonal factor models. Since all these values are standardised ($\mu = 0$, $\sigma^2 = 1$), it is assumed that a standard score on a data variable can be expressed as a linear combination of common, unique (specific) and error factor scores. i.e.

$$Z_{ik} = \sum_{j=1}^{m} a_{ij}F_{jk} + a_iS_{ik} + b_iE_{ik} \qquad (3)$$

where

Z_{ik} : Standard score for firm k on data variable i,

a_{ij} : factor loading for data variable i on common factor j,

a_i : factor loading for data variable i on the unique factor i,

b_i : factor loading for data variable i on error factor i,

F_{jk} : standard score for firm k on common factor j,

S_{ik} : standard score for company k on unique factor i,

E_{ik} : standard score for company k on error factor i.

While Z, F, S and E an all standardized, the "a" values are numeric constants ($a_{i1} \neq a_{i2} \neq a_{im}$, $F_{1k} \neq F_{2k} \neq F_{mk}$). A simultaneous representation of all values of i and k results in the following matrix:

$$Z = A_u F_u \qquad (4)$$

where

Z : matrix of data-variable scores,

A_u : matrix of factor loadings,

F_u : matrix of factor scores.

Factor extraction of Principal components

As indicated earlier, the application of factor analysis in this project is limited to the principal factor analysis method. The main reason for this choice was because this method takes care of total variance - commom variance + specific variance + error variance - as opposed to others which only evaluate common variance. Another reason for this choice was the fact that in recent years the principal factor method has become much more widely applied (see Comrey (1973)). This of course facilitates the standardization and future comparison of research results performed on common areas in different parts of the world. The approach is viz: using the correlation matrix as input, let R = Symmetric matrix of correlation coefficient. The ultimate objective is to determine two matrices - normalized eigenvector and eigenvalue - such that will yield R when multiplied together. In matrix algebra, a symmetric matrix (in this case R) can be diagonalized by premultiplying it by an orthogonal matrix (say B) and post multiplying it by the transpose of B to attain a diagonal matrix D. i.e.

$$BRB' = D \qquad (5)$$

As long as B is orthogonal, it follows that

$$BB' = B'B = 1$$

because sums of squares of rows and columns equal 1 and inner products of non identifical rows or columns equal to zero. Hence, if B is orthogonal, B' is the inverse of B and vice versa. Therefore equation (5) can be represented as

$B'BRB' = B'D$ (6)

Since $B'B$ = identity matrix, it drops out and equation (6) becomes

$RB' = B'D$ (7)

Multiplying both sides of equation (7) by $B \Rightarrow$

$RB'B = B'DB$ (8)

Again $B'B$ = identity matrix i.e. equation (8) is

$R = B'DB$ (9)

Equation (9) is then the required expression with an eigenvalue matrix D and eigenvector matrix B. In order to attain a factor matrix as required, let $\sqrt{D}$ be a diagonal matrix with elements in the diagonals that are the square roots of the corresponding elements in the diagonals of D.

Then equation (9) can be written as

$R = B'\sqrt{D}\sqrt{D}B = (B'\sqrt{D})\ (\sqrt{D}B) = AA'$ (10)

where

$(B'\sqrt{D}) = A$ or factor matrix

$(\sqrt{D}B) = A'$ or transpose of A.

The matrix obtained from factor analysing is a factor matrix such that $R = AA'$. Further more

$$\sum_{i=1}^{n} \sum_{j=i}^{n} (A_{ji})^2 = \lambda_i \qquad (11)$$

The proportion of variance extracted by factor i is $\dfrac{\lambda i}{\sum_{i=1}^{n} \lambda_i}$ (12)

where the denominator represents the total variance extracted because $\sum_{i=1}^{n} \lambda_i$ will equal the sum of the diagonal values of matrix R according to

matrix algebra. For each variable the R_{ii} (i = 1,2... n) gives the proportions of variance - h^2 - that are input to the analysis whose sum represents the total extractable common variance - communalities - in the matrix.

Let C = matrix of results, n = 13, m = 4

F_i (i = 1, 2, 3) = isolated common factors

h^2 = communality

C_{i1}, (i = 1,2... 10) = Proportion of attribute's variance accounted for by factor 1 (F_1)

C_{ij}, (i = 1,2... 10) = Proportion of attribute's variance accounted for by factor j (F_j) (j = 1,2,3)

C_{im}, (i = 1, 2...10) = Proportion of variance of attribute i contributed in the factor space (h_i^2)

$$\text{e.g. } C_{1m} = \sum_{j=1}^{m-1} (C_{ij})^2 = (-.69)^2 + (-.62)^2 + (-.09)^2 = 0.86 \qquad (13)$$

$$\text{In general, } C_{k1} = \sum_{k=1}^{n-3} \sum_{\ell=1}^{m-1} (C_{k\ell})^2$$

% total variance contributed by factor q (q = 1,3) is computed as follows:

$$C_{n-3q} = \frac{\sum_{q=1}^{m-1} \sum_{p=1}^{n-3} (C_{pq})^2}{n-3} * 100 \qquad (14)$$

where

C_{n-3q}: Percentage of variance among all the variables that is accounted for by the factors

$$\% \text{ common variance} = C_{n-1,v}, \quad (v=1,3) = \frac{C_{pq}}{\sum_{k=1}^{n-3} (C_{km})^2} *100 \qquad (14.1)$$

Where

$C_{n-1,v}$ = Variance among all the variables that is accounted for by factor v

$$\text{Eigenvalues } (\lambda_i) = C_{ni} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-3} (C_{ji})^2 \qquad (15)$$

In order to compute the relationship between the total variance and the coefficients in equation 3, equation 10 is further analysed. Since the matrix R (see eqn. 10) is symmetric, it follows that $R_u = R_u'$ because elements above the diagonal are equal to elements below. R_u = correlation matrix with diagonal equal to 1. From eqn. (10) $R_u = A_u A_u'$ (16)

The contribution made to the diagonal elements by the specific and error factors in eqn (16) can be calculated as follows:

square both sides of eqn. (3) ⇒

$$Z_{ik}^2 = a_{i1}^2 F_{1k}^2 + a_{i2}^2 F_{2k}^2 + \ldots + a_{im}^2 F_{mk}^2 + a_i^2 S_{ik}^2 + b_i^2 E_{ik}^2 + 2a_{i1}F_{1k}a_{i2}F_{2k}$$
$$+ \ldots + 2a_i S_{ik} b_i E_{ik} \qquad (17)$$

Adding both sides of eqn. (17) and dividing by N gives

$$\left(\frac{\sum_{k=1}^{n} Z_{ik}^2}{N}\right) = a_{i1}^2 \left(\frac{\sum_{k=1}^{n} F_{ik}^2}{N}\right)$$
$$+ a_{i2}^2 \left(\frac{\sum_{k=1}^{n} F_{2k}^2}{N}\right) + \ldots + a_{im}^2 \left(\frac{\sum_{k=1}^{n} F_{mk}^2}{N}\right)$$
$$+ a_i^2 \left(\frac{\sum_{k=1}^{n} S_{ik}^2}{N}\right) + b_i^2 \left(\frac{\sum_{k=1}^{n} E_{ik}^2}{N}\right)$$

$$+ 2a_{i1}a_{i2}\left[\frac{\sum_{k=1}^{n} F_{1k} F_{2k}}{N}\right] +\ldots+ 2a_i b_i \left[\frac{\sum_{k=1}^{n} S_{ik}E_{ik}}{N}\right] \qquad (18)$$

In eqn. (18), every term in parentheses is a variance of a standard score in either data variable or factor, and hence equals 1.0. Furthermore, in eqn. (18) every term in brackets is a correlation coefficient and is equal to zero because a correlation between uncorrelated factors is zero.

Simplifying eqn (18) ⇒

$$1 = a_{i1}^2 + a_{i2}^2 +\ldots+ a_{im}^2 + a_i^2 + b_i^2 \qquad (19)$$

Since the total variance involved is equal to 1 ($\mu = 0$, $\sigma^2 = 1$), the following equation can be evaluated in order to determine the unique factor.

i.e. total variance = 1 = common + unique + error variance

From eqn. (19) ⇒

$$S_{ik}^2 = h_i^2 + a_i^2 + b_i^2 = 1 \qquad (20)$$

where

h_i^2 = communality

a_i = coefficient of S_{ik}

b_i = coefficient of E_{ik}

Since there exists no correlation between the specific and error factors, the following relation between the uniqueness and its components can be found,

$$U_{ik}^2 = a_i^2 + b_i^2 \qquad (21)$$

Hence uniqueness = $1 - h_i^2$ (22)

While communality can be regarded as the proportion of a variable's total variance accounted for by the factors - sum of squared loadings for each variable; the uniqueness or unique variance indicates what magnitude of a variable is unrelated to the others, i.e. the degree to which the variance of a variable cannot be derived from the common factors F_i.

Transformed dimensions

After performing the oblique rotation, two matrices - pattern and structure - were produced. Even though the matrix clarified the dimensions, the sometimes plausible looking factors were not as evident as expected. The factor correlation matrix, however, revealed the independence of the ten factors. While most of the correlations were near zero, the highest correlation was - 0.208 between factor 3 and 5. The reference vector loadings that were the basis for factor interpretations are presented in order of importance. The hierarchical order which was determined by the amount of variance accounted for by each factor eliminated all variable loadings less than an absolute value of 0.300. Hence the assigned ranks are a function of the percentage of variance explained by each dimension. The identifications of factors were generally made from those variables with the highest factor loadings.

Factor I: Personnel - Quality

This factor defined human oriented operations such as control during manufacturing, education, experience, selection of personnel, time standard and degree of automation. Exports is the only variable with a high loading on this factor that does not appear to fit conceptually. Since this factor explains the highest percentage of variance, one might conclude that most firms attach the greatest importance to the variables loaded on it.

Factor II: Profitability

Factor two (which looks plausible) composed of rentability on operations, shareholders, assets and processes - weekly man hour in order to achieve improvement. This was identified as reflecting profit making operations.

Factor III: Material versus machinery

This identified material (predominantly raw) related variables - variability of input, losses on raw materials, age of equipment, specialized manpower and quality control. Perception of technology did not seem to fit conceptually this was attributed to a possible poor phrasing of the question.

Factor IV: Manufacturing

Factor 4 was composed of product type, subcontracting, manufacturing tolerance and department of research and development. Conceptually, research and development did not seem to belong to this group. This again could be attributed to the wording of the question.

Factor V: Set up versus cost

The variables which loaded on this factor were set up, frequency of set up, manpower costs on sales and cost price. This was identified as a cost/ set up factor which fitted well both operationally and logically.

Factor VI: Innovation versus production

This did not seem to portray any significant attributes apart from the variable's high loading. A restructuring of the question could eliminate the presence of such one variable factors. Hence its factor name is its variable's name.

Factor VII: Innovation versus work study

The diverse mixture of variables defining this factor made its interpretations very difficult. It was not conceptually easy to group the variables - measurement of ergonomics, incentives and innovation under one factor. This draw back could again be attributed to the structuring of the questions.

Factor VIII: Added value

This factor grouped variables relating to personnel experience - office and production departments - , added value and percentage of commercial success. Because the last variable loaded most, it was felt that identifying the factor by that name was more meaningful.

Factor IX: Relation-supplier

Factor 9 was made up of dependence on supplier and who the supplier was. This looked quite plausible as it grouped supplier related variables.

Factor X: Percentage - raw material

The 10th factor grouped product (fork shaped), percentage of raw materials on sales and lot size. Again this grouping did not seem to fit conceptually.

Observations

The overall factor interrelationship established earlier is low indicating a general presence of factor independence. Not surprising though is the correlation between factor three and five (0.208) which are both labelled physical process factors with "technology" and "set up" identified as unique attributes between both dimensions. A further analysis of the variable loadings has lead to the classification of the dimensions in two groups - major and minor - based on the total number of variables loaded on each dimension. Hence the major dimensions can be identified as factors 1, 2, 3, 4, 5, 8 while 7 and 9 can be regarded as minor dimensions.

The degree to which the factors account for or "explain" each of the variables is illustrated in Table 1. As shown, the factors extracted in this analysis account for "INTR001" better than they account for "PPHY004". The magnitude of the communality therefore is an invaluable index indicating how much of a variable is excluded after what it has in common with other variables has been extracted. The relatively low h^2 of "EFIN010", for example, suggests that it has little in common with the other variables included in the analysis. On the other hand, the comparatively high loadings of "PPHY010" and "EFIN006" show that they have much in common with manufacturing operations (considered collectively) that the factors represent. The most remarkable loading is that of "PPHY005" (0.908) indicating a very strong communality.

Table 1 Communality matrix (for variable description see Appendix I)

Variable	Communality	Variable	Communality	Variable	Communality
IDEN015	0.72898	PPHY012	0.59058	RHUM018	0.59789
IDEN016	0.41870	PPHY013	0.72438	RHUM021	0.57793
INTR001	0.63002	PPHY016	0.59568	RHUM026	0.49216
INTR002	0.61808	PPHY017	0.86097*	RHUM030	0.61364
INTR007	0.38444	PORG001	0.67012	EXTR001	0.46716
INTR009	0.49614	PORG004	0.55857	EXTR005	0.33837
INTR011	0.63031	PORG005	0.62450	EXTR007	0.41395
PPHY001	0.59490	PORG016	0.58709	EXTR009	0.35499
PPHY002	0.39064	PORG018	0.41051	EFIN004	0.58944
PPHY004	0.40600	PORG020	0.48693	EFIN005	0.50058
PPHY005	0.90834*	RHUM008	0.55118	EFIN006	0.81267*
PPHY007	0.63878	RHUM010	0.58110	EFIN010	0.31519
PPHY010	0.81489*	RHUM015	0.44963		

* Variables with very strong communalities

Homogeneous firms

In the preceding section, an analysis aimed at grouping variables based on their maximum distances - correlation coefficients - furnished us with independent clusters of variables. This section attempts to group companies based on minimizing the within - group company variance. The homogeneous enterprise grouping is done sequentially with the transpose of the raw data matrix serving as input (see Howard and Harris (1966)). The program like the one used to cluster the variables was run on a CDC Cyber 174.

Original dimensions

Based on the percentage of variance explained by the factors, ten dimensions were extracted from the 38 variables correlations matrix. The number of factors - ten - whose variances were between 4 to 15.5 percent was considered because factors outside that range did not reveal any meaningful results. The ten unrotated dimensions which serve only as a bench mark to the rotated solution, explained 67.3 percent of the total variance among the 38 variables.

The dimensions - factors - successively delineate the maximum amount of variance with each factor operating as a function of the total set of variables included in the analysis. The addition or subtraction of variables tends to fluctuate each factor. As soon as a rotation is performed, the fit of each dimension is maximized to a group of associated variables hence rendering each of these factors stable regardless of the variable modifications. An analysis of the oblique rotation follows.

Apparent attributes (see Table 2)

Group one which was generally identified as big production systems was pratically satisfactory except for company numbers 28 and 35 which did not fall in that category.

Group two - expensive furniture - was also conceptually meaningful. The companies which would not normally be classified in this group were numbers 9, 31, 76 and 77.

The third cluster identified as middle size furniture manufacturers was perfect from a practical point of view.

Latent attributes

Based on latent characteristics, column 10 (Table 2) indicated perfect groupings in groups 5, 8 and 9. In groups 1, 3 and 4, the misfits were company numbers 29, 41 and 36 respectively. Groups 6 and 7 were meaningfully grouped except for number 14 in group 6 which did not fit conceptually. The fourth group was made up of groups 2 and 10. In group 10, the first misfits were 67, 68 and 5 which empirically belonged to one of the perfect groups - 5, 8, 9. The second misfit was 55 which belonged to either group 6 or 7.

In the same manner, similar observations were made about the other columms of table 2 aimed at identifying the different company strata. This revealed that the bigger the company clusters, the more homogeneous their operations.

TABLE 2

CLUSTERING OF 41 COMPANIES

Company	CLUSTERS								
	2	3	4	5	6	7	8	9	10
63	1	1	1	5	5	5	5	5	5
67	2	3	3	1	6	6	7	7	10
68	2	3	3	3	3	7	7	7	7
69	2	3	3	3	3	7	7	7	10
73	2	3	3	3	3	7	7	7	7
02	2	2	4	4	4	4	4	4	4
03	2	3	3	3	3	7	8	8	7
04	2	3	3	3	3	7	8	8	8
05	2	2	2	2	2	7	7	7	10
06	2	3	3	3	3	7	8	8	8
07	1	1	1	1	6	6	6	6	6
08	1	1	1	5	5	3	3	3	3
09	2	2	4	4	4	4	4	4	4
11	2	3	3	3	3	7	8	8	8
14	1	1	1	1	1	1	6	6	6
18	2	3	3	3	3	7	8	8	8
20	1	1	1	5	5	5	5	5	5
22	1	1	1	5	5	5	5	5	5
24	1	1	1	5	5	3	3	3	3
28	1	1	1	1	6	6	6	6	10
29	2	2	2	1	1	1	1	1	1
30	2	2	2	2	2	2	2	2	2
31	2	2	2	2	2	7	2	2	7
32	2	2	2	2	2	2	2	2	7
34	1	2	2	2	2	3	3	3	3
35	1	1	1	1	6	6	7	7	10
36	2	2	4	4	4	4	4	4	4
41	2	3	3	3	3	3	3	3	3
45	1	1	1	1	1	1	1	1	1
46	2	3	3	3	3	7	7	7	10
47	2	3	3	1	1	1	1	1	1
53	2	3	3	3	3	7	7	7	7
55	1	1	1	1	6	6	7	7	10
56	2	3	3	3	3	7	7	7	7
57	1	2	2	2	2	2	2	2	5
58	1	1	1	1	6	3	6	6	6
59	1	1	1	5	5	5	5	5	5
60	1	2	4	4	4	3	3	3	3
62	1	1	1	1	1	1	1	1	1
76	2	2	2	2	2	2	2	2	2
77	2	2	2	2	2	2	2	9	9

Conclusions

This research has attempted to present the theoretical and practical concepts of multidimensional applications particularly of interest to industrial engineers. Its importance in a given production set up could be summarized viz:

(1) Parameter identification that could be suitably used to describe a given industry.

(2) Quantification and classification of the above variables in order of importance.

(3) Stratification of companies in a given industry.

The advent of numerical taxonomy has presented yet another essential tool in data analysis to managers and analysts. This technique was applied here because of the availability of both the tools and the data. Since its findings as presented above are very interesting, one cannot hesitate to say that these methods will be here for a while.

APPENDIX 1

CLUSTER REPRESENTATION OF VARIABLES

Appendix 1 Analysis of Variable clusters (Level= 0.2)

Cluster	Variable	Description
1	INTR008	Consultation
	*PORG004	Cost price
2	*PPHY010	Set up frequency
	RHUM002	# of employees - CEGEP
	RHUM025	Description of tasks
3	* PORG009	Sales forecast
	PORG003	Production plan
	EXTR001	Quality control
4	*INTR013	Subcontract for others
	PPHY013	Production method vs sectorial innovation
	PPHY014	Production equipment vs sectorial innovation
5	*PORG016	Control during manufacturing
	RHUM022	Office staff with < 1 year experience
6	*RHUM018	Office staff with > 15 years experience
	RHUM023	Man power turn over

Appendix 1 (continued)

7	RHUM014	#employees with university education
	*INTR009	Business with which supplier
	RHUM020	Office staff with < 10 and > 6 years
8	*PORG020	Department of R & D
	EXTR006	Life of products
9	RHUM016	#employees with commercial education
	*RHUM021	Office staff with < 5 and > 1 year
10	*PPHY016	% costs of sales employees
	RHUM029	Employees belong to union
11	RHUM011	#employees half specialized
	EXTR003	Added value grouped
	*EXTR009	Added value in dollars
12	PPHY003	Handling
	*PPHY017	% raw material cost on sales
13	PPHY006	Process control
	*RHUM008	#employees with experience < 1 year and > 5 years
14	PORG008	Environment
	INTR003	Mean tolerances
	*PPHY001	Degree of automation

Appendix 1 (continued)

	EXTR004	New products
	EXTR011	% sales increase at production level
15	*INTR001	Dependence on supplier
	INTR010	Structured documentation centre
	RHUM027	Personnel evaluation
16	PORG019	Management by coho
	*RHUM010	Specialized man power
17	*IDEN016	Product vs prop. #1
	EXTR002	Product complexity
18	*PORG018	Minimum batch size
	RHUM017	#employees with < commercial education
19	INTR012	% of process subcontracted
	EXTR010	Product vs sectorial level
	*EFIN004	Rentability of operation year 1.
20	*PPHY007	Perception of technology
	PPHY011	State of technology
21	PPHY009	Standardized handling
	*PORG001	Time standard

Appendix 1 (continued)

22	INTR005	Loss on raw material grouped
	*INTR011	Loss on raw materials in %
23	RHUM013	Man power turn over
	*RHUM030	Personnel well rewarded
24	RHUM003	# of employees with secondary education
	*PPHY005	Set up
	RHUM006	# employees with < 20 years and > 11 years experience
	RHUM019	Office staff with < 15 years and > 11 years experience
25	RHUM005	# employees with > 20 years experience
	PORG002	Use of computer
	*RHUM026	Personnel selection
26	PORG006	Job security
	RHUM007	# employees with < 10 and > 6 years experience
	*RHUM015	# employees with CEGEP education
	EFIN002	Director of sales (excluding manager)
27	*PPHY004	Manufacturing tolerance
	RHUM004	# employees with less than secondary education
	RHUM009	# employees with less 1 year experience
	RHUM012	Non specialized man power

Appendix 1 (continued)

28	RHUM031	Regular meetings with management?
	*EFIN009	Prototype designs
29	PORG011	Quality control
	*INTR002	Input variability
	RHUM024	Formation and perfection
30	INTR004	Technical & scientific information
	PPHY008	Understanding of process
	*PPHY012	Weekly man hour to improve process
31	*EFIN005	Shareholder rentability year 1
	EFIN003	Financial rentability year 1
	EFIN006	Assets rentability year 1
32	PORG007	Reduction of costs
	*EXTR007	Innovation
33	*EXTR005	Export
	EFIN001	Sales in Quebec
34	*INTR007	Sub contracts
	PORG010	Control
	PPHY015	Equipment handling vs sectorial level
	RHUM001	# employees from university

Appendix 1 (continued)

35	PORG005	Ergonomic measurements
	*EFIN008	Objectif #2
36	*PPHY002	Age of equipment
	PORG012	Management
	RHUM028	Mean age of production workers
37	*IDEN015	Product vs type #1
	EXTR008	Degree of product transformation
38	INTR006	Manufacturing under licence and patent
	*EFIN010	% commercial success

* Most representative variables selected for factor analysis.

Latent attributes 35

Conclusions 37

Appendix 1 - Cluster representation of variables 38

References 45

References

BROWN, R.V., 1970, Do managers find decision theory useful?, Havard Business Review, p. 79.

CATTELL, R.B., 1972, The basis of recognition and interpretation of factors, Educational and psychological measurement 22: pp. 667-97.

COMREY, A.L., 1973, A first course in factor analysis, Academic Press New York and London.

DEVENDRA, S., 1975, The generalized distance measurement of technology, Tech. Fore, Soc. Change 9, pp. 289-300.

DIXON, W.J., 1975, BMDP Biomedical Computer Programs, (Berkeley, Los Angeles, London, University of California, Press).

GODFREY, E.P., FIEDLER, F.E., and HALL, D.M., 1958, BOARDS, Management and Company Success, Danville, Ill., (Interstate Printers and Publishers).

GREEN, P.E., 1978, Analyzing Multivariate Data, (Dryden Press Hinsdale, Illinois).

GREEN, P.E., and RAO, V.R, 1972, Applied Multidimensional Scaling: A Comparaison of Approaches and Algorithms (New York H, R & W).

HARMAN, H.H., 1976, Modern Factor Analysis, rev. ed. Chicago, (University of Chicago Press).

HARPER, R. et al., 1950, The application of multiple factor analysis to industrial test data. Brit. Journ. App.Physics 1,1.

HOWARD, N., and HARRIS, B., 1966, A hierarchical grouping, IBM 360/65 FORTRAN IV program (University of Pennsylvania Computer Centre).

JOHNSON, S.C. 1967, Hierarchical clustering schemes, psychometrika XXXII, pp. 241-254.

KENDALL, M.G., 1968, A course in multivariate analysis (Hafner New York).

KRUMBEIN, W.C., and IMBRIE, J., 1963, Stratigraphic factor maps, American Association of Petroleum Geologists, 47 pp. 698-701

LAINJO, B.D., 1980, Analysis of manufacturing systems unpublished masters thesis, Ecole Polytechnique, University of Montreal.

LAWLIS, G.F., and CHATFIELD, D., 1974, Multivariate approaches for the behavioral sciences: A brief text (Texas Tech. Press).

MEYER, S.L., 1975, Data analysis for scientists and engineers (John Wiley and Sons, Inc.)

RUMMEL, R.J., 1970, Applied factor analysis (Evanston: Northwestern University Press).

SOKAL, R.R., 1966, Numerical taxonomy, Scientific American, 215, pp. 106-16.

VILLENEUVE, L., and CHARRON, G., 1977, Mesure de la technologie: revue et projections (Rapport technique EP77-R-24, Montreal, Ecole Polytechnique).