10. Factor Analysis in Marketing

Jagdish N. Sheth
University of Illinois

Douglas J. Tigert
University of Toronto

Introduction

The principle of parismony is common to all scientific theory. It states that a construct or a network of constructs (model) should be simpler than the data upon which it is based. Take, for example, a two-construct model which states that store patronage is a function of the social class to which the patron belongs. Although only two constructs are involved in the model, the data required to test it would be quite large: first, we must obtain a sample of observations to provide distributions of each of the constructs. And secondly, we must define and measure a large number of manifest indicators of each of the constructs. In the case of store patronage, these may for example be: (a) number of trips to the store in a predefined time interval, (b) total number of trips to all stores in the same period, (c) amount of expenditures at the store in that time interval, and (d) self-reported extent of store preference. In the same manner, the social class of the patron could be reflected by his income, occupation, education, and dwelling unit.

From an inductive and empirical observation point of view, a construct is an aggregate abstraction of a large variety of observed phenomena based on the presence of common characteristics among them. In the case of social class, for example, what we do is to abstract, simplify and collate such diverse phenomena as income, education, occupation and dwelling unit. Indeed, it could very well be that we, in fact, started out with many more manifest phenomena and finally ended up with these four indicators. This process of abstraction and delimitation is fundamental to any scientific theory (Howard and Sheth, 1969). The uniqueness of an indicator or the fallibility in its observation (error) is discarded in favor of commonality of that indicator with other indicators. In short, a construct is an abstraction of reality in which the correlations among indicators are retained or summarized for the sake of parsimony.

Factor analysis is one of the most powerful multivariate methods that enables a researcher to simplify and summarize a large data matrix into a much smaller one without appreciable loss of information. It is primarily concerned with the resolution of a set of manifest variables (indicators) linearly in terms of new constructs or factors, and hence, its principal aim is to attain scientific parsimony or economy of description. As Massy has put it: "Factor analysis is basically a method of reducing a set of data into a more compact form while throwing certain properties of the data into bold relief" (Massy, 1964, p. 291).

Factor Analysis as a Multivariate Method

The place of factor analysis in the statistical family of multivariate methods is excellently described by Kendall (1950, pp. 60-61). He classifies the study of statistical relationships into two broad categories: (i) analysis of dependence, and (ii) analysis of interdependence. The former includes analysis of variance and regression analysis, while the latter includes various types of correlation analysis and contingency analysis. The important distinction between the two categories is that the techniques under analysis of dependence all require the designation of one or more of the total variables in analysis as dependent, while the techniques under analysis of interdependence focus attention upon relationships among the total set of variables without singling out any of them for special consideration.

Kendall considers factor analysis in the category of analysis of interdependence. However, he also suggests that, unlike other methods in the same category, (e.g., multidimensional scaling), factor analysis is far on the scale towards analysis of dependence. This is because factor analysis creates linear dependence between observed or manifest variables and hypothetical or derived factors, as will be described later. Indeed, some researchers consider that all the conventional least squares methods of multiple rectilinear prediction from a set of independent variables to a set of dependent variables (multiple regression, discriminant analysis, and canonical correlations) are special cases of the more generalized formulation of factor analytic theory (Horst, 1965, p. 21). It is true, however, that none of the *manifest* variables is explicitly treated as dependent on other manifest variables, and to that extent it is classified as a method for analysis of interdependence among manifest variables.

In factor analysis the principal aim-- parsimony without appreciable loss of information-- is achieved by linking factor analysis to some powerful algebraic and geometric concepts. Suppose we have a data matrix X of sixe 10 × 1000 in which each element x_{ji} represents a consumer i's evaluation of a product on the j^{th} attribute (i = 1, 2, 3, ..., n; N = 1000 and n = 10 in this particular example). For any one of the characteristics x_j, the system (x_{j1}, x_{j2}, x_{j3}, 111, x_{jN}) of N real numbers can be considered a point in an N-dimensional space. However, by considering each system as a vector, we may *simplify* this configuration to ten dimensions although it has to be regarded as imbedded in an N-space (N=1000). In general, the configuration of n vectors may be regarded as in an n-dimensional space which is imbedded in the original N-space. We have already achieved considerable parsimony by using simple geometrical notions that reduce a 1000 dimensional space to a much smaller 10-dimensional space. Hence, it simplifies the description of the original configuration to the minimum dimensions possible.

This simplification in factor analysis is based on the amount of linear dependence that exists in a data system. If some variables are linearly dependent upon other variables, it is possible to describe them in terms of these other variables themselves. Then by retaining only that subset of variables which are linearly independent, we may always explicitly describe others by stating their dependence on this subset. Hopefully, the number of such linearly independent variables will be

considerably smaller in a large set of data so that by retaining them only, simplification would result.

It can be shown (Kendall, 1961) that the minimum number of dimensions to which a configuration in a data matrix X of size n × N (n < N) can be reduced is its rank r (r ≤ n). These r dimensions then form a factor space on which the n-vectors can be projected.

A second and intimately related function of factor analysis is still unresolved. One of the characteristics of factor analysis, and some of the basic mathematical concepts underlying it, is that if one finds such a matrix of summarization, then it is also true that an infinite number of such sets exist, each of which is equally useful in reproducing or accounting for the information in the original data matrix. In other words, there is indeterminacy inherent in factor analysis. This has necessitated the second objective of factor analysis, *factorial invariance*, by the procedures of rotations based on principles of simple structure. It entails choosing decision rules for selecting only one of an infinite number of sets possible in the process of simplification.

To summarize then, factor analysis reduces a data matrix X of size n × N to a factor matrix F of size r × N in which r is generally much smaller than n, which simplifies the data description. The original data matrix X is, however, linearly dependent upon F and the coefficients of this dependence are presented in another matrix A of size n × r. Then a given observation x_{ji} can be approximated with minimum loss of information by

$$x_{ji} \sim \hat{x}_{ji} = a_{j1}F_{1i} + a_{j2}F_{2i} + \dots + a_{jr}F_{ri} = \sum_{m=1}^{r} a_{jm}F_{mi} \quad (1)$$

Then, in matrix notation,

$$X \sim \hat{X} = AF \quad (2)$$

Matrix A is called the ***factor loadings*** matrix, and matrix F is called the *factor scores* matrix.

Description of Factor Analysis

Factor analysis simplifies a data matrix by summarizing each variable's variance in a small number of factors. This is done by creating a linear dependence[1] of the variable on a set of common factors and a unique factor. Hence,

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jr}F_r + d_jU_j \quad (3)$$

or,

$$Z_j = \sum_{m=1}^{r} a_{jm}F_m + d_jU_j$$

The total variance of a variable Z_j (in standard scores) is summarized by a set of common factors (F_m) and by a unique factor (U_j). That portion of a variable's variance which covaries with other variables in the data is classified among the common factors and the balance is retained in the unique factor.

The total variance of a variable j is simply the average of squared standard scores of a sample of N observations, i.e., $S_j^2 = \sum_{i=1}^{N} Z_{ji}^2/N$

Then, based on equation (3)

$$S_j^2 = \Sigma Z_{ji}^2/N = \sum_{m=1}^{r} a_{jm}^2[\Sigma F_{mi}^2/N] + d_j^2 \Sigma U_{ji}^2/N$$

$$+ 2\sum_{m<p=1}^{r} a_{jm}a_{jp}[\Sigma F_{mi}F_{pi}/N] + 2d_j \sum_{m=1}^{r} a_{jm}[\Sigma F_{mi}U_{ji}/N] \quad (4)$$

In the above equation, it is assumed that all the variables, including factors, are in standard form. Then

$$S_j^2 = 1 = \sum_{m=1}^{r} a_{jm}^2 + d_j^2 + 2\sum_{m<p=1}^{r} a_{jm}a_{jp}r_{FmFp} + 2d_j \sum_{m=1}^{r} a_{jm}r_{FmUj} \quad (5)$$

As will be shown later, the unique factor (U_j) is always uncorrelated with the common factors ($r_{FmUj} = 0$) and, therefore, the last term on the right side of equation (5) vanishes. If we also assume that the common factors are uncorrelated among themselves ($r_{FmFp} = 0$, $m \neq p$) as would be expected if they represent orthogonal dimensions of a space, then one more term in equation (5) vanishes. This leaves the more simplified expression.

$$S_j^2 = 1 = \sum_{m=1}^{r} a_{jm}^2 + d_j^2 = a_{j1}^2 + a_{j2}^2 + \ldots + a_{jr}^2 + d_j^2 \quad (6)$$

Thus, it will be seen that each a_{jm}^2 represents a portion of total variance in Z_j which is summarized in factor Fm. In a similar manner, the same factor Fm will summarize part of the variances of other variables. The total contribution of common factors toward summarizing the variance of Z_j, called h_j^2, then equals

$$h_j^2 = \sum_{m=1}^{r} a_{jm}^2 = a_{j1}^2 + a_{j2}^2 + \ldots + a_{jr}^2 \quad (7)$$

This is technically referred to as the *communality* of Z_j; it represents the portion of total variance in Z_j that is common to other variables. Later it will be shown that each factor loading (a_{jm}) represents a product-moment correlation between the variable j and factor Fm if the data are presented in standard form and factors are orthogonal to one another. It will be then seen that a_{jm}^2 represents a squared multiple R in the regression sense. Therefore, each h_j^2 can be thought of as a squared multiple R between a variable j and the common factors. The total communality, h^2, of all the variables in a data matrix is then

$$h^2 = \sum_{j=1}^{n} h_j^2 = \sum_{j=1}^{n}\sum_{m=1}^{r} a_{jm}^2 \quad (8)$$

Obviously, the unique variance (d_j^2) of a variable Z_j is ($1-h_j^2$). However, this variance itself can be thought of as consisting of two components: stable or true unique variance and error variance. In other words,

$$1 - h_j^2 = d_j^2 = g_j^2 + e_j^2 \quad (9)$$

Therefore,

$$S_j^2 = 1 = h_j^2 + g_j^2 + e_j^2 = \sum_{m=1}^{r} a_{jm}^2 + d_j^2$$

Having shown that the total variance of a variable is summarized in a set of common factors and a unique factor, we may write equation (3) for all the n variables as follows:

$$\begin{aligned} Z_1 &= a_{11}F_1 + a_{12}F_2 + \ldots, + a_{1r}F_r + d_1U_1 \\ Z_2 &= a_{21}F_1 + a_{22}F_2 + \ldots, + a_{2r}F_r + \quad d_2U_2 \\ Z_n &= a_{n1}F_1 + a_{n2}F_2 + \ldots, + a_{nr}F_r + \quad d_nU_n \end{aligned} \quad (10)$$

Thus, a data system of n variables is summarized in r common factors and n unique factors, or a total of (r+n) factors. But (r+n) > n, which seems contradictory to the rule of parsimony fundamental to factor analysis. This apparent contradiction stems from the assumption about the equivalence of communality with the total variance of a variable or unity. The unique factors are not derived from the resolution of a data matrix but are simply calculated as residuals from the difference between unity and communality. The usual procedure followed is either to include total variance of the variables in factor analysis, or only that part of it which represents communality. If the former is used, then the assumption is that total variance, including the unique portion, will be approximated by the factor matrix. In that case, the model is as follows:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \ldots, + a_{jr}F_r \quad (11)$$

and

$$S_j^2 = 1 = \sum_{m=1}^{r} a_{jm}^2$$

This is commonly referred to as *principal components analysis* as opposed to *classical factor analysis*, the latter being represented by equation (3).

On the other hand, if the second course of action is adopted, then the common factors will approximately reproduce the communality of the variable ($h_j^2 < 1$), and a unique factor will be necessitated to account for total variance.

While the two models (equations [3] and [11]) are distinctly different, the procedures outlined later in the paper are essentially the same in the derivation of factor loadings and factor scores matrices. The only difference is that while total variance ($S_j^2 = 1$) is retained in principal components analysis, an estimate of communality is substituted for total variance ($S_j^2 = h_j^2$) in the data matrix before factoring in classical factor analysis.[2]

Since matrix notation compactly represents the total data system, we would like to express the above notions of decomposition of variance in matrix notation. Let a data matrix of standard scores be called Z which has Zji elements (j = 1, 2, ..., n; i = 1, 2, ..., N) representing individual i's score on variable j. Let F be a factor scores matrix with F_{mi} elements (m = 1, 2, ..., 4; i = i, 2, ..., N) representing individual i's score on factor Fm. Finally, let A be an n × r matrix with a_{jm} elements (j = 1, 2, ..., n; m = 1, 2, ..., r) representing the linear weights between Z and F. Then equation (3) can be expressed as

$$Z = AF + DU \quad (12)$$

in which AF represents common factor space and DU represents unique factor space. It will be recalled that DU is a diagonal matrix.

If only common factors are involved, as would be the case in principal components analysis, then we may write equation (12) as

$$Z = AF \quad (13)$$

which represents equation (11) in matrix notation. We will be primarily concerned with the common factors only and, hence, ignore the DU matrix hereafter.

If both sides of equation (13) are post-multiplied with F' and divided by the scalar N, we obtain

$$ZF'/N = A(FF'/N) \quad (14)$$

Remembering that all the scores are in standard form, the left side of (14) represents correlations between manifest variable scores and factor scores. Let this be called $S = ZF'/N$.

On the right side of (14) the expression in parentheses represents correlations among factor scores. Let us express this as $\phi = FF'/N$. Then

$$S = A\phi \quad (15)$$

If the factors are uncorrelated, as would most likely be the case for the initial results, then ϕ is an identity matrix (I) and, therefore

$$S = AI = A \quad (16)$$

This clearly shows that the factor loadings matrix A represents the correlations between manifest variable scores and factor scores.

Let R represent the correlations among the manifest variables. R can be obtained from the standard scores data matrix Z as follows:

$$R = ZZ'/N = A(FF'/N)A' = A\phi A' \quad (17)$$

If the factors are uncorrelated ($\phi = I$), then the correlations among manifest variables can be reproduced from the factor loadings matrix itself:

$$R = AIA' = AA' \quad (18)$$

This is important to remember because it states that the product of factor loadings matrix will reproduce the original correlations.

Finally, the factor scores matrix F can be easily obtained *if* (and this is important to remember) A is a square matrix. Then, there exists an inverse of A, and from the basic equation (13) we may write

$$F = A^{-1}Z \quad (19)$$

However, A is usually a vertical matrix of $n \times r$ ($r < n$) and, therefore, it does not have an inverse. The only case where A will have an inverse is when there are as many factors (including trivial ones) as there are variables, in which case it will be a square matrix.

It is, however, still possible to calculate the factor scores matrix (F) by following multiple regression principles. Consider each factor as a *dependent* variable to be estimated from a linear combination of manifest scores. Then the following is true (in standard form):

$$F_m = \beta_{m1}Z_1 + \beta_{m2}Z_2 + \ldots, \beta_{mn}Z_n \quad (20)$$

and for any individual i this becomes

$$F_{mi} = \beta_{m1}Z_{1i} + \beta_{m2}Z_{2i} + \ldots, + \beta_{mn}Z_{ni} \quad (21)$$

The vector B representing beta weights β_{mj} then can be obtained as follows:

$$B = R_{12}R_{11}^{-1} \quad (22)$$

Where R_{11} represents correlations among independent variables and R_{12} represents correlations between a dependent variable and each of the independent variables. If we follow the logic of equation (20), R_{11} is the correlation matrix among manifest variables (R) and R_{12} is the factor loading matrix (A), representing correlations between factors and manifest variables. Then

$$B = A'R^{-1}$$

and

$$F = A'R^{-1}Z \quad (23)$$

If the factors are correlated, then $S = A$ will be substituted in place of A in equation (23) and the relationship will still hold.

To summarize, a *factor* (Fm) is a linear combination of manifest variables (Z_j). Hence,

$$F_m = \sum_{j=1}^{n} \beta_{mj}Z_j \quad (24)$$

A *factor score* (F_{mi}) is the linear combination of an individual i's manifest scores:

$$F_{mi} = \sum_{j=1} \beta_{mj}Z_{ji}; \text{ or } F = BZ = A'R^{-1}Z \quad (25)$$

A *factor loading* is the correlation between a factor score and a manifest score:

$$a_{jm} = \sum_{i=1}^{N} Z_{ji}F_{im}/N; \text{ or } A = ZF' \quad (26)$$

In the Appendix the derivations of the factor loadings and factor scores matrices are described in detail.

Rotation of Factors

As described in equation (2), a data matrix X can be approximated by $\hat{X}$, which is obtained as a product of the factor loadings matrix (A) and the factor scores matrix (F). However, if a matrix can be expressed as a product of two other matrices, we know that it can also be expressed as the product of an infinite number of pairs of matrices. This results in the indeterminacy of factor results, and can be shown as follows:

Let us post-multiply the factor loadings matrix A by any square orthonormal matrix T ($T^{-1} = T'$ and, therefore, $TT' = I$). Then we obtain another matrix B such that

$$B = AT \quad (27)$$

We also multiply the factor scores matrix F with the same orthonormal matrix T and obtain another matrix V such that

$$V = T'F$$

Then $\hat{X} = BV = ATT'F$. (28)

Since there is an infinite number of T matrices of order r, there would be an infinite number of pairs of matrices satisfying the basic linear postulate in factor analysis.

In the above illustration, we assumed that T is an orthonormal matrix in which the product of any pair of vectors is zero. This is the orthogonality property that is retained in deriving the B and V matrices. It is, however, not necessary that the transformation matrix T be orthonormal. Any square basic matrix P of order r would be sufficient to produce an infinite number of pairs of B and V matrices. Thus

$$B = AP$$

and

$$V = P^{-1}F \text{ so that} \tag{29}$$

$$\hat{X} = BV = APP^{-1}F$$

If the square transformation matrix is not orthonormal, the product of any pair of vectors will be non-zero. This type of transformation is referred to as an oblique transformation, and it will be described a little later.

Given the indeterminacy in factor analysis, Thurstone (1947) suggested some judgmental guidelines for choosing one and only one of these infinite transformations, based again on the fundamental principle of parsimony. His guidelines are called *principles of simple structure* and they are summarized below from Harman (1967, p. 98) in terms of five decision rules.

1) Each row of the factor loadings matrix ($A_{n \times r}$) should have at least one zero.

2) If there are r common factors, each column vector of $A(A_{j1}, A_{j2}, \ldots, A_{jr})$ should have at least r zeroes.

3) For every pair of columns in the factor loadings matrix (e.g., A_{j1} & A_{j2}) there should be several variables whose entries vanish in one column but not in the other.

4) For every pair of columns in the factor loadings matrix, a large proportion of the variables should have zero or near zero entries in both columns when there are four or more factors.

5) For every pair of columns in the factor loadings matrix there should be only a small number of variables with non-zero entries in both columns.

Choosing a single pair of matrices with the use of these decision rules is commonly referred to as the rotation problem because it implies rotating factor axes from a given position to the desirable or ideal position. Also underlying this rotation is the rule of parsimony, but it is not intuitively as simple as it was in the case of factoring the data matrix. It may be helpful to think that the rule of parsimony is applied in factoring by focusing attention on the dimensions or axes of the configuration, whereas in rotation it is applied by focusing on the configuration of vectors of variables. Thus, in rotation some structure is sought in the clustering of configurations by moving axes from position to position. Indeed, the goal in rotation is to move from a completely chaotic projection of a configuration to a completely meaningful projection in which each variable tends to attain unit complexity.

Prior to computers, principles of simple structure were attained by manual and mechanical devices that gave adequate and satisfactory configurations but not unique ones. Today, a variety of analytical rotations are available, each of which gives an invariant factor matrix. All of them use some or all of the principles of simple structure. The types of analytical rotations are broadly classified as *orthogonal* and *oblique* rotations.

In orthogonal rotations the factors are kept uncorrelated in the process of rotating them in the factor space. Hence they maintain the property of orthogonality, and the transformation matrix T which moves them from one position to another is an orthonormal (and therefore an orthogonal) matrix. This is equivalent to using equation (27). On the other hand, in oblique rotations, the factor axes are deliberately passed through the most obvious clusterings of variables, which do not necessarily fall at right angles. This means that the orthogonality present in the initial factoring is given up for the sake of, hopefully, making more sense out of clusters of variables. Hence, equation (29) is used for transforming the initial factor solution to the rotated factor solution. This results in the factors themselves being correlated, which is summarized in ϕ in the notation of equation (15).

There are two major orthogonal type rotations: quartimax and varimax. In the quartimax (and its obverse, quartimin) approach, the rule of parsimony underlying the simple structure principles is achieved by attempting to pass a factor axis through a point (variable) so that it can be fully described in terms of a single factor; the focus is on simplifying rows (variables) of a factor loadings matrix. However, it is found that this sometimes leads to the situation in which all the variables are simplified in terms of the *same* factor with the consequence that a general factor emerges. To avoid this, and to present a different perspective, varimax rotation attempts to simplify columns (factors) of the factor loadings matrix. Here, an attempt is made to describe a factor in terms of as few variables as possible. This results in transforming the factor loadings of a factor (column) as close to either unity or zero as possible when the input is a matrix of standardized data and, therefore, correlations.

Oblique rotations are also based on the same rule of parsimony underlying the principles of simple structure. Three variations of orthogonal quartimax rotations are commonly used in oblique rotations. These are called oblimax, oblimin and quartimin. Also, an oblique version of the more preferred orthogonal varimax rotation--covarimin--is now available. Finally, a compromise of both covarimin and quartimin called the biquartimin oblique rotation is computerized for easy use.

The mathematical derivatives of both orthogonal and oblique rotations are beyond the scope of this paper. The reader is referred to Harman (1967) for an elegant treatment on this aspect of factor analysis.

Applications in Marketing

Among all multivariate methods, factor analysis is probably the most widely used technique in marketing except possibly multiple regression (Sheth, 1968a, 1969). However, factor analysis has only recently been applied in marketing (Sheth and Armstrong, 1969). One way to examine the use of factor analysis in marketing is to look at the emphasis placed on a given aspect of factor analysis in a study. A search of the marketing literature revealed three major emphases: (1) extraction of factors and investigation of their meaning, (2) structural relationship among variables as projected on factor axes, and (3) derivation of factor scores for further investigation.

Extraction of Factors. Factor analysis has been used in marketing to extract underlying factors from simple single questions on products and brands, for example, asking a respondent to rate his preference for the product or brand. Similar ratings are obtained on a variety of products and brands from a sample of respondents to create a data matrix which is then factor analyzed. Primarily, two objectives are sought in extracting meaningful (rotated) factors.

First, it is hoped that the factors will provide insights as to the causal links that underlie the relative ratings of various products and brands. The causal factors are presumed to determine the clustering of products on various factors, and knowing them is considered equivalent to adding surplus meaning to the data. A typical study is reported by Stoetzel on the liquor preferences of the French people (Stoetzel, 1960). A national probability sample of 1442 respondents was asked to rank order nine types of liquors popularly consumed in France. The original raw data were converted to standard scores to derive a 9 × 9 product-moment correlation matrix. This was factor analyzed by centroid factor analysis and the factors were graphically rotated based on Thurstone's principles of simple structure. Three meaningful factors were extracted which, when combined together, explained a large part of the total variance in the data matrix.

Stoetzel used these three factors to explore the underlying causal characteristics of liquors that presumably determined liquor preferences in France. He labeled these factors as (1) sweetness-strength, (2) low-high price, and (3) regional popularity. The labeling of the third factor was determined after examining the average rankings of various liquors in samples of major regions of France. The factor loadings matrix is reproduced in Table 1. According to Stoetzel, "The major principle of liquor preference in France is the distinction between sweet and strong liquors. The second motivating element is price, which can be understood by remembering that liquor is both an expensive commodity and an item of conspicuous consumption. Except in the case of the two most popular and least expensive items (Rum and Marc), this second factor plays a much smaller role in producing preference judgments. The third factor is concerned with the sociological, and primarily the regional, variability of the judgments." (p. 11).

Table 1. Stoetzel Study on Liquor Preferences

TYPES OF LIQUOR	FACTORS I	II	III
Liquers	.64	.02	.16
Kirsch	.50	-.06	-.10
Mirabelle	.46	-.24	-.19
Rum	.17	.74	.97[1]
Marc	-.29	.66	-.39
Whiskey	-.29	-.08	.09
Calvados	-.49	.20	-.04
Cognac	-.52	-.03	.42
Armagnac	-.60	-.17	.14

[1]This figure is certainly too high and should not exceed the maximum of .64.

Presumably, factor analysis of simple rankings revealed considerable information to the researchers as to why the French prefer various types of liquor. However, unless there is external validating evidence from the sample respondents, it is possible that this surplus meaning added to the data based on subjective judgments may prove misleading.

A second objective in extracting meaningful factors from a data matrix based on a single simple question is to cluster, and thereby segment, a variety of objects, including television programs, magazines and products (Swanson, 1967; Wells, 1967; Wells, Banks, and Tigert, 1967; Tigert, 1969). A typical study of this clustering and segmentation approach is presented by Wells in factor analyzing the frequency of readership of 30 magazines (Wells and Sheth, 1974). The 30 × 30 matrix of correlations among magazines was factor analyzed by the principal components method and factors were orthogonally rotated by the varimax procedure. A total of 10 factors were extracted explaining about 69 percent of total variance. Table 2 reproduces the rotated factor loadings of 30 magazines on these 10 factors. The magnitudes of the factor loadings clearly indicate systematic clustering of magazines of distinct types and specializations.

Factor 1 has high loadings on *Car and Driver*, *Road and Track*, *Motor Trend* and *Hot Rod*. This means that respondents who say they read *Car and Driver* also tend to say they read the other magazines that correlate highly with factor 1. In other words, these four magazines form a group based on some degree of common audience. Factor 2 has relatively high loadings on *Fortune*, *Forbes*, *Time* and *Business Week*. Again, the interpretation is that magazines in this group have more audience overlap with each other than with other magazines. The inference is that they, like the first group, represent some common core of interests. Factor 3 represents *Field and Stream*, *Outdoor Life* and *Sports Afield*; and factor 4 represents *Farm Journal*, *Successful Farming* and *Progressive Farmer*. Similar clusterings of homogeneous magazine types is evident among all the 10 factors.

Table 2. Factor Analysis of Magazine Readership

VARIABLE DESCRIPTION	VAR. NO.	1	2	3	4	5	6	7	8	9	10
		(Decimals omitted)									
Bus. Week	1	05	51	06	-10	02	03	-15	26	-06	43
Life	2	07	20	04	-06	73	01	-09	08	09	10
New Yorker	3	00	37	04	-11	05	-05	21	53	-07	19
Time	4	-00	65	02	-06	31	-01	-07	07	08	12
Newsweek	5	03	10	09	-09	20	04	05	20	11	68
U.S. News & World Rpt.	6	-03	12	00	12	07	03	14	-02	-06	72
Sat. Review	7	-04	-03	-03	09	17	09	26	69	-11	10
Look	8	-6	-05	09	-01	71	09	13	00	08	20
Sat. Ev. Post	9	-05	03	07	03	74	08	11	03	10	-01
Forbes	10	-07	64	-00	02	-06	12	15	-07	-03	09
Argosy	11	17	00	17	-01	12	04	01	01	75	07
Atl. Mthly.	12	-02	08	02	-02	07	02	78	07	03	13
Car & Driver	13	78	04	01	-04	04	-10	-02	-04	-06	01
Fld. & Stream	14	05	04	84	04	09	10	-00	-02	14	04
Farm. Jrnl.	15	-01	-03	05	87	-02	-02	-04	-02	-00	09
Fortune	16	03	72	-02	-00	01	-01	19	14	07	-03
Harpers	17	08	15	-01	-03	07	-09	75	15	-02	-01
Mech. Illus.	18	26	10	15	-01	09	72	-01	-03	09	-01
Pop. Mech.	19	16	03	19	00	03	82	00	04	05	05
Pop. Sci.	20	02	00	03	-04	09	81	-07	10	04	05
Outdoor Life	21	02	01	80	04	01	13	01	01	13	-02
Prog. Farmer	22	-04	07	07	61	07	05	02	03	-13	-25
Reader's Dgst.	23	-24	12	-14	15	21	18	-07	-08	36	34
Road & Track	24	77	03	10	-03	06	04	-01	10	01	-07
Sci. Amer.	25	07	08	01	-00	-05	06	-04	74	14	-01
Succ. Frmng.	26	-01	-11	07	75	-08	-06	-03	-01	12	17
Sports Afield	27	08	-03	78	10	08	08	-00	01	07	06
True	28	06	05	23	-03	13	10	02	04	75	-06
Hot Rod	29	68	-07	-01	-01	-04	16	05	-03	13	01
Motor Trend	30	77	-03	05	03	01	20	04	03	12	04

It is, however, worth noting that *Reader's Digest* did not have a high loading on any one factor. Instead, its positive loadings were divided among the news group (factor 10), general reading group (factor 5), men's fiction group (factor 9) and farming group (factor 4). This result implies that *Reader's Digest* has an appeal that spreads broadly across readers of at least four specialized magazine types. It also has small negative loadings on the sports car group (factor 1) and hunting and fishing group (factor 3), implying that readers of these magazine types tend not to read *Reader's Digest* (Wells and Sheth, 1974, pp. 16-19).

Structural Analysis of Factors. In this type of application of factor analysis in marketing, a data matrix is factored to observe the structural relationship among the variables in terms of their projections on factors. Thus the emphasis is more on the configuration of variables than on axes. Once again, there are two distinct objectives in this approach.

First, interest may be centered on the distribution of a set of variables on various factors from one study to another study, either on the same sample of respondents or otherwise (Clevenger, Lazier and Clark, 1965). A typical study in this area is reported by Sheth and Ring (1968) in which they attempted to compare the clustering of 14 attributes (13 characteristics plus one overall rating) across three product categories of instant breakfast, dietary products, and milk additives. In particular, they were interested in comparing the structure of attribute correlates of general attitude as manifested by the overall rating.

A sample of 954 housewives was asked to rate the most popular brand in each product category on the same set of 14 characteristics which were selected based on a pilot study. A total of six factors were extracted which accounted for more than 90 percent of variance. The factor analysis used the principal components method, and factors were rotated by both varimax and quarimax rotations with very comparable results. In all the three product categories, the overall rating was primarily a unidimensional concept (in this sample) in view of the fact that it loaded heavily on only one factor. Only with instant breakfast did it tend to be shared by a second factor that explained 7 percent of the total variance, as compared with 61 percent explained by the first factor. Table 3 reproduces the factors that summarized the variance of the overall rating. Sheth and Ring concluded that the most dominant attribute correlates (e.g., taste and flavor) are the same across the three product categories, but the other attributes vary in their relationship with the overall rating from one product class to the other.

A second objective in focusing on the configuration of variables on factors may be to search for the most promising variables that correlate with one single criterion variable so that they may be used in further analysis. For example, Twedt (1952) was interested in finding the determinants of advertising readership. He obtained a recall-based readership measure on 122 advertisements of 1/4 page or more in size from a single issue of a trade magazine (*American Builder*). Then, a total of 34 variables were developed based on mechanical and content aspects of these advertisements. These included size, color, illustration, number of product benefits, Flesch readability scores, etc. From the original 35 × 35 variables correlation matrix, Twedt isolated 19 predictor variables that had significant correlations with the readership criterion. Then he factor analyzed the 19 variables using the centroid method, obtained six meaningful factors that summarized a large percent of total variance, and graphically rotated them using Thurstone's principles of simple structure. Table 4 reproduces these rotated factors. By inspection, he felt that there were three predictor variables that seemed to load highly on these factors and which also summarized a large part of the variance in the readership variable. These were: (1) size of the advertisement, (2) number of colors and (3) square inches of illustration.

Twedt used these three variables as predictors and the readership variable as the criterion in a multiple regression, and obtained a multiple R of .75. He attributed this success to using factor analysis for exploring relationships among variables. When an additional six variables were included in multiple regression, the multiple R jumped to only .79, which suggests that factor analysis had adequately isolated three salient determinants of readership. Twedt then used only these three variables

in determining readership of five other types of magazines and found that they consistently had good predictive power, since multiple R ranged from .58 to .80.

Table 3. Dimensions of General Attitude and Its Correlates*

	MILK ADDITIVE	DIETARY PRODUCT	INSTANT BREAKFAST	
	Factor	Factor	Factors	
	I	I	I	II
Overall Liking	.816	.812	.781	.260
Variance in Liking summarized by factors	67%	66%	61%	7%
Description	Attribute	Correlates		
Delicious Tasting	.834	.877	.799	
Has Real Flavor	.767	.805	.803	
Easy to Use	.721	.129	.377	
Very Good For Snack	.682	..509	.377	
Good Buy For Money	.524	.397	.354	
Provides Lots of Energy	.312	.230	.227	.787
Poor Substitute	.216	.420	.347	
Good Source of Protein				.775
Very Filling				.678
Low in Price				
Low in Calories				
Does not Dissolve Easily				
Somewhat Nutritious				

*Only factors summarizing five percent or more of the variance in the overall rating are reproduced in this table. Also, only loadings greater than .200 are reproduced here.

Table 4. Selection of Variables from Factor Analysis (Twedt Study)

	Factors						
	I	II	III	IV	V	VI	h^2
Readership	64	35	28	18	16	09	68
1. Square Inches of Illustration	51	48					
2. Pictures of Use	51	23					
3. Number of Colors	49	-07					
4. Size of Ad	18	69					
5. Largest Type			62				
6. Number of Copy Blocks				68			
7. Number of Product Identifications				64			
8. Number of Product Facts	20	37	-24	28			
9. Surround			61		76		
10. Previous Schedule	42	46				47	
11. Brad-Vern Schedules	21	37	-25			34	
12-19 Not included here; Total variables = 35							

Regression Analysis Results

Ad Readership = f(1,3 and 4); R = .76
Ad Readership = f(1 through 11); R = .79

Derivation of Factor Scores. A third set of applications of factor analysis in marketing has focused attention on factor scores. Since factor scores are linear combinations of manifest scores, they serve as useful summary variables for further investigation. Furthermore, for orthogonal factors, the factor scores usually are uncorrelated, thus removing collinearity from the manifest data. Unlike the previous applications, factors are treated as the dependent variables and manifest variables as the independent variables. Three objectives have been sought in the emphasis on factor scores.

First, if a large number of items are rated by a sample of respondents, it is difficult and sometimes impossible to use all items for further investigation, such as determinants of a specific behavior. It is then desirable to simplify without much loss of information the large number of observations on the respondents by reducing manifest scores to factor scores. A typical study is reported by Bass, Pessemier and Tigert (1969). From a sample of 344 housewives, magazine reading habits were obtained based on a six-point scale ranging from "Never read" to "read almost every issue." In addition, each respondent answered an AIO (activity-interest-and-opinion) scale consisting of 300 six-point items, 8 demographic questions, a personality test and an occupational choice test. In short, there were more than 500 observed or manifest variables, and it would have been virtually impossible to regress magazine readership on all of them simultaneously. The three sets of independent variables, namely AIO, personality scores and occupational choice scores were separately factor analyzed; this yielded 14, 8 and 5 factors, respectively. This, added to the 8 demographic variables, resulted in 35 total independent variables, a manageable set of variables in multiple regression.

A second objective in focusing on factor scores is primarily to reduce both collinearity in the data by factoring raw data and converting raw scores to factor scores. A good example of this application is from Campbell's study on the determinants of the evoked set of brands for toothpaste and detergents. Campbell (1968) hypothesized that the size of the evoked set (number of brands the buyer considers as choice alternatives within a product class) is a function of socioeconomic, demographic and psychological variables. He collected information on 12 such variables; and prior to regressing the evoked set on them, he factor analyzed them using principal components analysis and rotated the resulting six factors by the orthogonal varimax method. The six factors summarized 69 percent of total variance and are reproduced in Table 5. Campbell then used stepwise regression in which the evoked set was treated as the dependent variable and each of the six factors as an independent variable. He found that in the case of detergents, the six factors explained 35 percent of the variance in the evoked set and in the case of toothpaste, 27 percent. Furthermore, in both cases, factor 4 (price-brand loyalty) was the most significant predictor, it alone explaining 30 and 23 percent of the variance in the evoked sets of detergent and toothpaste.

Finally, the focus has been placed on factor scores due to the belief that they contain substantive meaning just as factors do. In other words, factor scores as linear combinations of manifest scores are values of an underlying construct. For example, Sheth (1968, 1970) has treated factor scores, derived from a binary sequence of purchases of a brand over trials, as brand loyalty scores because brand loyalty is an unobservable construct to be inferred from the number and pattern of purchases of the brand within a specified time interval.

Table 5. Factor Analysis of Determinants of Evoked Set (Campbell Study)

VARIABLE	FACTORS 1	2	3	4	5	6
Relative S. E. Status	.82					
Total Income	.72					
Level of Education	.66					
Size of Family		.74				
Frequency of Purchase		.84				
Self-Confidence			.79			
Importance of Product			.70			
Importance of Price				.82		
Degree of Brand Loyalty				-.81		
No. of Brands Aware					.61	
Perceived Risk in New Products					.73	
Age						.88
Percent variance summarized	14	13	12	11	9	9 = 69%

STEP-WISE MULTIPLE REGRESSION OF EVOKED SET ON SIX FACTORS

TOOTHPASTE	percent variance explained	DETERGENT	percent variance explained
Factor 4	23		30
plus 3	25	plus 6	33
plus 1	26	plus 3	34
plus 2	27	plus 1	35
plus 6	27	plus 2	35
plus 5	27	plus 5	35

Precautions in Using Factor Analysis

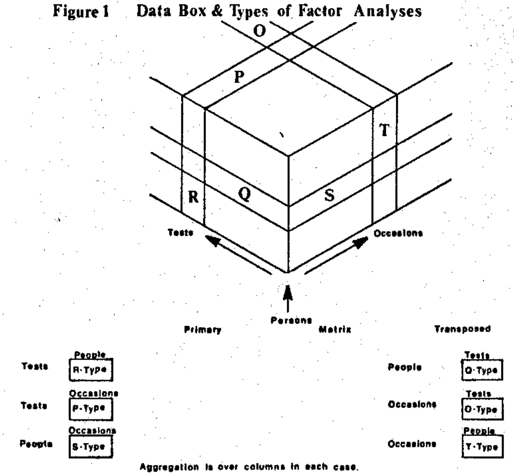
Like most other complex statistical methods, some precautions should be maintained in the use of factor analysis in marketing. Several researchers in marketing have noted that the variety and complexity inherent in factor analysis have tended to infatuate persons with the technique itself, and consequently, it is sometimes misused (Ramond, 1963; Ekebald and Stasch, 1967; Ehrenberg, 1968, Sheth and Armstrong, 1969). Below some of the important precautions in using factor analysis are pointed out.

(1) Factor analysis begins with a data matrix and not with a matrix of correlations among variables (Horst, 1965). As pointed out in the Appendix, a correlation matrix is only an intermediate step in factor analyzing data, although it is a useful and necessary step for obtaining factors by the use of the eigenvalue-eigenvector theorem. Indeed, it can easily be shown that factor analysis can be done on other types of square symmetric matrices (cross-products and covariance) that are also meaningful from the research viewpoint. The raw data are at first converted to standard scores (from which the correlation matrix is derived) to make the results of factor analysis independent of origins and scales of measurement. It then follows that in instances where all the variables have the same natural origin and are measured on the same scale, there is no necessity to standardize the data matrix and subsequently factor analyze a correlation matrix. In fact, it is pointed out that by standardizing such data, we

unnecessarily lose two vital bits of information, namely, the level and the dispersion of a variable (Sheth, 1969).

(2) Factor analysis is not limited to the covariances of variables. In marketing, we typically collect data on three distinct dimensions: entities (people, consumers, stores, brands), tests (variables, attributes, characteristics) and occasions (various time periods or trials). Thus, it is a data box which permits six distinct types of factor analyses (Cattell, 1952). These are called R-type (correlations among tests over entities at a given occasion), Q-type (correlations among entities over tests at a given occasion), P-type (correlations among tests over occasions for an entity), O-type (correlations among occasions over tests for an entity), S-type (correlations among entities over occasions for a given test), and T-type (correlations among occasions over entities for a given test). A graphical description of these six types of factor analyses is provided in Figure 1.

Figure 1 Data Box & Types of Factor Analyses



Some researchers create special two-dimensional matrices by stacking the third dimension in a certain manner (Burt, 1941, Sheth, 1966). For example, if we have measures of purchases from a sample of consumers (entities), on several brands (tests) over trials or time-intervals (occasions), it is possible to consider each trial as multichotomous in which a brand becomes one of several categories. Then a two-dimensional matrix (trials × consumers) is readily created for factor analysis. Finally, there is the possibility of factor analyzing a data box, as it is, by the procedures of three-mode factor analysis (Tucker, 1967).

(3) Factor analysis tends to overemphasize the factor loadings matrix as the only output, because it summarizes the correlation matrix, and the latter is often considered as the original input data. This overemphasis tempts the researcher to interpret his data solely in terms of factor loadings. However, the factor loadings may sometimes be misleading when, based on them, labels are attached to factors. For example, two variables may be uncorrelated and yet have a high loading on a factor. This can arise if each variable has a loading of .50 on the first factor, and one has a loading of +.50 on the second factor and the other has -.50 on it. This makes them uncorrelated variables (Nunnally, 1967, Ehrenberg, 1968), but the researcher may label the first factor based on the combination of these two variables.

In a similar manner, when there are only a small number of variables, factor loadings tend to be moderately high although the correlations among the variables are small; once again, defining factors based on these loadings may prove misleading.

Finally, when numerous combinations of a small set of variables are created and treated as additional variables (e.g., total purchases, average purchases, total purchases on deals, average purchases on deals, relative total purchases of two brands, etc.), additional factors will usually be obtained which may be interpreted in terms of artificially created variables.

(4) Without simultaneous validity and reliability of the observed data, it is possible to extract meaningful factors from randomly generated data (Armstrong and Soelberg, 1968). Although this is true of any statistical method (hence tests of significance are devised to reject the null hypothesis of chance behavior), factor analysis seems particularly vulnerable to adding meaning to random data because of the indeterminacy of factor solutions. Armstrong and Soelberg (1968) have suggested three procedures to build reliability into factor results: (1) split the sample, run separate factor analyses on each of them and compare the two results; if the data are truly random, there should be no congruence between the results; (2) have a prior theory which is as structured as possible so it can act as a benchmark; and (3) for small samples, do a Monte Carlo simulation using matched simulated data for comparisons.

A number of analytical methods have been proposed recently for validating various factor results. The reader is referred to Harman (1967, Chapter 12) and Cattell (1966, chapter 10), since these are beyond the scope of this paper.

APPENDIX

It is often stated that what a researcher factor analyzes is a matrix of correlations among variables. This is a misconception. While it is true that a correlation matrix usually is the input to factor analysis, it should be remembered that correlations are themselves obtained from the raw data. The starting point for any statistical method, including factor analysis, is the observed data matrix. Also any transformation of the raw data does not add more information, although it is possible to lose some information.

We shall therefore discuss the derivation of factor loadings and factor scores in terms of the original data. Take a data matrix X of size $n \times N$ in which an element x_{ji} represents individual i's score on j^{th} variable ($j = 1, 2, \ldots, n$; $i = 1, 2, \ldots, N$). Knowing that all the observed data are fallible in some manner or other, it is possible to postulate that the data matrix X is a sum of two matrices, one of which represents true, systematic or stable measures and the other random, unsystematic measures.

Thus, each $x_{ji} = \hat{x}_{ji} + e_{ji}$. Therefore,

$$X_{n\times N} = \hat{X}_{n\times N} + E_{n\times N} \tag{30}$$

Factor analysis is concerned with the problem of determining the $\hat{X}$ matrix in such a way that it will closely resemble the original data matrix X, and still have a much lower dimensionality (rank) than the X matrix. To achieve this, a powerful theorem of linear algebra called the basic structure of a matrix is utilized.

Any complete n × N rectangular matrix X (n < N) can be resolved as the product of three other matrices possessing some special characteristics. Thus

$$X_{n\times N} = U_{n\times n}\Lambda_{n\times N}F_{N\times N} \tag{31}$$

where U is an orthonormal matrix ($U^{-1} = U'$, $UU' = I$), Λ is a diagonal matrix containing λ_n values in the upper left section and zeroes elsewhere, and F is another orthonormal matrix ($F^{-1} = F'$, $FF' = I$).

Then the approximate matrix $\hat{X}$ of rank r (r × n) is constructed simply by using only some of the rows and columns of the three matrices (U, Λ and F). Furthermore, this is proven to be the best approximation to the original data matrix X in the least squares sense (Eckart and Young, 1936).

To choose the first r rows and columns, let us partition each of the three matrices as follows:

$$U_{n\times n} = [U_{n\times r}, U_{n\times e}]$$

$$\Lambda_{n\times N} = [\Lambda_{r\times r}, \Lambda_{e\times e}] \tag{32}$$

$$F_{N\times N} = [F_{r\times N}, F_{e\times N}]$$

and

$$n = r + e$$

Then

$$\hat{X}_{n\times N} = U_{n\times r}\Lambda_{r\times r}F_{r\times N} \tag{33}$$

and

$$E_{n\times N} = U_{n\times e}\Lambda_{e\times e}F_{e\times N}$$

Hence

$$X = \hat{X} + E = [U_{n\times r}\Lambda_{r\times r}F_{r\times N}] + [U_{n\times e}\Lambda_{e\times e}F_{e\times N}] \tag{34}$$

A factor loadings matrix $A_{n\times r}$ having a_{jm} elements is defined as the product of $U_{n\times r}\Lambda_{r\times r}$. A factor scores matrix F having F_{mi} elements is defined by $F_{r\times N}$. Therefore,

$$X \sim \hat{X} = AF \tag{35}$$

From equation (30) it is seen that

$$E = X - \hat{X} \tag{36}$$

The diagonals of EE' contain the sums of squares of the residual elements for each of the n variables, and these can be shown to be a function of the diagonals (trace) of $\Lambda^2_{e\times e}$. It is this relationship which provides the basis for considering $\hat{X}$ as the best approximation to X in the least squares sense.

The resolution of a data matrix X into the product of three matrices, U, Λ, and F is accomplished by first post-multiplying X with its transpose X'. Then

$$XX' = (U\Lambda F)(F'\Lambda U') = U\Lambda^2U' \tag{37}$$

remembering that $FF' = I$.

The left side of equation (37) is also the cross-products matrix that contains the sums of squares and sums of cross-products of the raw elements x_{ji} in X. This is important to realize, because if the x_{ji} elements were transformed to deviation scores $(x_{ji} - \bar{x}_j)$ and summarized in a matrix Y with y_{ji} elements $(y_{ji} = x_{ji} - \bar{x}_j)$, then the cross products matrix YY' would be equivalent to a covariance matrix except for dividing elements by the scalar N. Finally, if the raw data were converted to standard scores $(x_{ji} - \bar{x}_j/s_j)$ and summarized in a matrix Z with z_{ji} elements $(z_{ji} = x_{ji} - \bar{x}/s_j$ or $y_{ji}/s_j)$, then the cross products matrix ZZ' would represent a matrix of intercorrelations among variables except for dividing elements by the scalar N.

From this it can be seen that factor analyzing a correlation matrix is itself a special case of a more generalized approach.

In equation (37), the cross-products matrix XX' is a square symmetric matrix, and therefore it has a characteristic function with characteristic roots and vectors. This function will be stated later when discussing a correlation matrix. Finding the characteristic roots and corresponding vectors of XX' (or YY' and ZZ') can be mathematically shown to be equivalent to maximizing the variance summarized in the approximate matrix $\hat{X}$ (e.g., Harman, 1967, pp. 137-143).

In equation (37), the U matrix on the right side contains the characteristic vectors and the Λ^2 matrix contains the characteristic roots of XX'. It can be seen immediately that: (1) the vectors of X are the same as the characteristic vectors of XX' and (2) the roots of X are the square roots of the characteristic roots of XX'.

Thus, we have derived two of the three product matrices (U and Λ) which form the basic structure of X as defined in equation (31). It is then easy to obtain F by the following equation:

$$F = \Lambda^{-1}U'X \tag{38}$$

remembering that $U^{-1} = U'$.

Then, partitioning each of these three matrices of equation (32) and employing equation (33), we obtain the X and E matrices.[3]

Factoring a Correlation Matrix

As mentioned before, the correlation matrix R is the cross-products of the standard score matrix Z whose elements are then divided by the scalar N. Thus,

$$R = ZZ'/N$$

which is equivalent to

$$R = N^{-1/2}ZZ'N^{-1/2} \tag{39}$$

in which $N^{-1/2}$ is an n × n scalar matrix with elements $1/\sqrt{N}$ in the diagonals. Then the data matrix in whose basic structure we are interested when correlations are used is

$$N^{-1/2}Z = N^{-1/2}U\Lambda F. \tag{40}$$

Multiplying it with its transpose to resolve it into the product of three matrices we obtain

$$N^{-1/2}ZZ'N^{-1/2} = R = N^{-1/2}U\Lambda FF'\Lambda U'N^{-1/2} \tag{41}$$

The characteristic roots of R are summarized in the Λ^2 matrix and the corresponding vectors in $N^{-1/2}U$. Since R is a square symmetric matrix, it contains positive characteristic roots corresponding to its rank, and also has the matching vectors.

The characteristic roots and vectors are obtained as follows: The basic task in factor analysis is to determine an axis in the n-dimensional test space along which the variance is a maximum, then a second axis, orthogonal to the first, which accounts for as much of the remaining variance as possible, a third axis, orthogonal to both the first and the second, and so on. Each new orthogonal axis accounts for a smaller proportion of the original variance.

This task of finding axes that successively summarize maximum variance in a data matrix is shown to be equivalent to finding the characteristic roots and vectors of R (Harman, 1967). It entails solving a set of n homogeneous equations written as

$$U_1(1-\lambda) + U_2r_{12} + \ldots, + U_nr_{1n} = 0$$

$$U_1r_{21} + U_2(1-\lambda) + \ldots U_nr_{2m} = 0 \qquad (42)$$

$$\vdots$$

$$U_1r_{n1} + U_2r_{n2} + \ldots U_n(1-\lambda) = 0$$

Equation (42) can be written in matrix form as $(R - \lambda_m I)U_m = 0$, m = 1, 2, 3, ..., n, where I is an identity matrix, 0 is a null vector, and R is the correlation matrix. There are n nontrivial solutions and each U_m represents the coefficients for converting manifest scores into factor scores.

The nontrivial solutions are obtained by the characteristic equation of a square matrix, which in the above case means that the determinant of the coefficient of U is zero: $|R - \lambda I| = 0$. This gives the λ_m values which can then be used in equation (43), where in the beginning λ_1 is placed in the diagonals. Finally, knowing $r_{jk}(k,k = 1, 2, 3, \ldots, n)$ and m enables us to solve for U_m in equation (33).

In the above case, note that unity was placed in the diagonal elements of R, i.e., $r_{jj} = 1$. This means that the total variance ($S_j^2 = 1$) was to be factored in terms of common factors only, and therefore, technically it is principal components analysis. If the unities are replaced with communalities ($h_j^2 \leq 1$), and the same procedure is followed, we would be using the classical factor model as described in equation (3).

From equation (41) the factor loadings matrix of a correlation matrix R and the standard scores matrix $N^{-1/2}Z$ equals

$$A = N^{-1/2}U\Lambda$$

and the factor scores matrix

$$F = \Lambda^{-1}U'N^{1/2} \cdot N^{-1/2}Z = \Lambda^{-1}U'Z$$

Then

$$N^{-1/2}Z = AF = N^{-1/2}U\Lambda\Lambda^{-1}U'Z$$

Thus, the factor loadings matrix is composed of the characteristic vectors of R multiplied by the square root of the characteristic roots and further multiplied by the scalar $1/\sqrt{N}$.

NOTES

1. It is not mandatory that the dependence be linear. Recently, nonlinear factor analysis has been investigated by several researchers. See McDonald (1962).

2. This may seem overly simplified to those who are familiar with the controversy among psychologists on the issue of communality. Researchers like Burt (1941), Horst (1965) and Nunnally (1967) would consider both the types of factor analyses as equivalent. On the other hand, other researchers, notably Lawley & Maxwell (1963), Kendall (1961), and Harman (1967) would emphasize the difference. Kendall has even sharpened the difference by considering principal components analysis as an inductive, empirical approach, and classical factor analysis as a hypothesis-testing or deductive approach. In marketing, researchers are primarily concerned with empirical data due to lack of a theory and, therefore, we think more along the lines of principal components analysis.

3. One more step is usually involved to rescale the factor scores and factor loadings matrices in order that the results can be compared across different sample sizes. This is because the orthonormal property of $F(F^{-1} = F'; FF' = I)$ is based on unit length vectors of N, and therefore is directly related to the size of the sample used in a study. Then two studies, otherwise identical, but differing in sample sizes, will not be directly comparable. See Sheth (1969) for standardization procedures.

REFERENCES

Armstrong, J. S. and Soelberg, P., "On the Interpretation of Factor Analysis," *Psychological Bulletin*, 70 (1968), 361-364.

Barban, A. M. and Grunbaum, W. F., "A Factor Analytic Study of Negro and White Responses to Advertising Stimuli," *Journal of Applied Psychology*, 49, 274-279.

Bass, F. M., Pessemier, E. A. and Tigert, D. J., "A Taxonomy of Magazine Readership Applied to Problems in Marketing Strategy and Media Selection," *Journal of Business*, 42 (1969), 337-363.

Burt, C., *The Factors of Mind*, New York: MacMillan, 1941.

Campbell, B. M., *The Existence and Determinants of Evoked Set in Brand Choice Behavior*, (Unpublished Ph.D. Dissertation, Columbia University, 1968).

Cattell, R. B., "The Three Basic Factor Analytic Research Designs--Their Interrelationships and Derivatives," *Psychological Bulletin*, 49 (1952), 499-520.

Cattell, R. B., (ed.) *Handbook of Multivariate Experimental Psychology*, Chicago: Rand McNally, 1966.

Clevenger, T., Lazier, G. A., and Clark, M. L., "Measurement of Corporate Images by the Semantic Differential," *Journal of Marketing Research*, 2 (1965), 80-82.

Eckart, C. and Young, G., "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, 1 (1936), 211-218.

Ehrenberg, A. S. C., "On Methods: The Factor Analytic Search for Program Types," *Journal of Advertising Research*, 8 (1968), 55-70.

Ekebald, F. A. and Stasch, S. F., "Criteria in Factor Analysis," *Journal of Advertising Research*, 7 (1967), 48-57.

Harman, H., *Modern Factor Analysis*, 2nd Edition, Chicago: University of Chicago Press, 1967.

Horst, P., *Factor Analysis of Data Matrices*, New York: Holt, Rinehart and Winston, 1965.

Howard, J. A., and Sheth, J. N., *The Theory of Buyer Behavior*, New York: Wiley, 1969.

Kendall, M. G., "Factor Analysis as a Statistical Technique," *Journal of the Royal Statistical Society*, 12 (1950), 60-73.

--------, *A Course in Multivariate Analysis*, London: Griffin, 1961.

Lawley, D. N., and Maxwell, A. E., *Factor Analysis as a Statistical Method*, London: Butterworth, 1963.

Massy, W. F., "Applying Factor Analysis to a Specific Marketing Problem," in Greyson, S. A., ed., *Toward Scientific Marketing*, Chicago: American Marketing Association, 1964, 291-307.

McDonald, R. P., "A General Approach to Nonlinear Factor Analysis," *Psychometrika*, 27 (1962), 397-415.
Nunnally, J. C., *Psychometric Theory*, New York: McGraw-Hill, 1967.
Ramond, C. K., "Factor Analysis: When to Use It," in A. Shuchman, ed., *Scientific Decision Making in Business*, New York: Holt, Rinehart, and Winston, 1963.
Sheth, J. N., *A Behavioral and Quantitative Investigation of Brand Loyalty*, (Unpublished Ph.D. Dissertation, University of Pittsburgh, 1966), Chapter 2.
--------, "A Factor Analytic Model of Brand Loyalty," *Journal of Marketing Research*, 4 (1968), 395-404.
--------, "Estimating Parameters Using Factor Analysis," *Journal of the American Statistical Association*, 64 (1969), 808-822.
--------, "Measurement of Multidimensional Brand Loyalty of a Consumer," *Journal of Marketing Research*, 7 (1970), 348-354. (b)
--------, "Multivariate Analysis in Marketing," *Journal of Advertising Research*, 10 (1970), 29-40 (a).
Sheth, J. N. and Armstrong, J. S., "Factor Analysis of Marketing Data: A Critical Evaluation," AMA Educator's Conference, 1969.
Sheth, J. N. and Ring, W., "Correlates of General Attitudes," July, 1968. Working paper, School of Business Administration, University of Illinois, Urbana.
Stoetzel, J., "A Factor Analysis of the Liquor Preferences of French Consumers," *Journal of Advertising Research*, 1 (December, 1960), 7-11.
Swanson, C. E., "The Frequency Structure of Television Magazines," *Journal of Advertising Research*, 7 (June, 1967), 8-14.
Thurstone, L. L., *Multiple Factor Analysis*, Chicago: University of Chicago Press, 1947.
Tigert, D. J., "A Psychographic Profile of Magazine Audiences: An Investigation of a Medium's Climate," presented at the AMA Workshop on Consumer Behavior, 1969.
Tucker, L. R., "Some Mathematical Notes in Three Mode Factor Analysis," *Psychometrika*, 31 (1966), 279-311.
Twedt, D. W., "A Multiple Factor Analysis of Advertising Readership," *Journal of Applied Psychology*, 36 (1952), 207-215.
Wells, W. D., "Patterns of Consumer Behavior," AMA June 1967 Conference, pp. 134-138.
--------, "Backward Segmentation," in J. Arndt, ed., *Insights into Consumer Behavior*, Boston: Allyn and Bacon, 1968.
Wells, W. D., Banks, S., and Tigert, D. J., "Order in the Data," AMA Winter 1967 Conference, pp. 263-266.
Wells, W. D. and Sheth, J. N., "Factor Analysis in Marketing Research," in R. Ferber (ed.) *Handbook of Marketing Research*, New York: McGraw-Hill, 1974.
Wells, W. D. and Tigert, D. J., "Activities, Interests and Opinions," *Journal of Advertising Research*, 2 (August, 1971), 27-35.

10a. Discussion and Comments

Benjamin F. King
University of Washington

I should like to make a few cautionary remarks on the use and interpretation of factor analytic models. I prefer to discuss factor analysis as a model, or a family of models, rather than a *technique*; if the Sheth and Tigert paper has any shortcomings, they are (1) a certain muddiness about the distinction among variants of the general factor model, and (2) the danger of leaving the uninformed with the impression that the only empirical solution to the factor model is the principal components approach, with or without estimated communalities on the diagonal. After all these years, the paper by Anderson and Rubin (1956) is still the definitive description of factor analysis from the statistical viewpoint.

We all know there is no unique solution to the basic factor analytic equation. To achieve identification, one has to impose strong restrictions on the system, restrictions that are likely to be arbitrary and of questionable origin. It is one thing for a researcher to talk about factors and their loadings, as if they were real entities, when he has a well-developed prior theory that dictates a particular factor pattern and the restrictions necessary for obtaining a solution. It is not so awful that the factors are unobservable; what is deplorable is for a researcher to give names to unseen factors that he has extracted from an ensemble of variables when:

1) he does not fully understand, and he has not weighed carefully the assumptions underlying his mathematical method of obtaining the factor solution;

2) he has no carefully formulated hypothesis to begin with; and

3) he has no intention of replicating his experiment to check out any hypothetical factor patterns that may have been suggested by the present data.

There are some statisticians who have no use for factor analysis at all. Some people in marketing research might place Ehrenberg (1968) in that category. In the discussion of a 1950 paper on the topic by Babington-Smith, G. A. Barnard said, "When the proposal for a discussion on factor analysis was made, I expressed the hope that someone would denounce the whole method. It seems that this was the reason why I was asked to join in the discussion."

To be fair to Barnard, he excludes from the subject of factor analysis the method of principal components, and after he heard Sir Cyril Burt say that nearly all psychologists "do give pride of place to the method of principal axes" (which, as Sheth and Tigert observe, is not exactly the