

FACTOR ANALYTIC APPROACH IN THE STUDY OF INDUSTRIAL PRODUCTION DATA

Sivakumar .K¹, Srinivasan .M.R²

¹Department of Mathematics, Sathyabama University, Jeppiaar Nagar, Rajiv Gandhi Road, Chennai, India

²Department of Statistics, University of Madras, Chennai, India

E-mail : ¹shree_eashean@yahoo.com

Abstract

Industrial Production data is the key to study the growth and development of a nation. The data needs to be disseminated before formulating any models for prediction. Factor Analysis a key tool in multivariate statistics and has been used to determine the latent factors that best explain the data. In addition Principal Component Analysis framework has been used for analyzing the data with respect to the contribution of variability by each of the components. The multiplicative models has been examined for its initial strength of the relationship for the production data and validation of the same has been discussed. The data has been processed for the removal of outliers, if any.

Key words : Production Data, Principal Component Analysis, Communalities, Component Matrix, Rotated Component Matrix and Outliers.

I. INTRODUCTION

Factor Analysis is a statistical approach that can be used to analyze interrelationships among a large number of variables and to explain these variables in terms of common underlying dimensions (factors). The main applications of factor analytic techniques are to reduce the number of variables and to detect structure in the relationships among the variables. Therefore, factor analysis is applied as a data reduction or structure detection method (the term factor analysis was first introduced by Thurston, 1931). This paper presents a study of Cosmetics and Toiletries which is one of the industrial production data. Using Principal Component Analysis generaliability of finding for whole set of data been discussed. If the absolute value of factor scores are greater than 3 an outlier is identified. A less costly alternative approach split sample validation technique been applied and compared to the results.

II. METHODOLOGY

If variables are standardized, the model may be represented, as the unique factors are uncorrelated with each other and common factors as

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \dots + A_{im}F_m + V_iU_i$$

X_i = ⁱth Standardized Variable

A_{ij} = Standardized Variable

F = Common Factor

V_i = Standardized regression co-efficient of variable i on unique factor i .

U_i = the unique factor for variable i

M = No. of common factor.

Once observations are collected, the correlation matrix is computed and examined to determine factor analysis appropriate. If nearly all the correlations are small, there is probably not much point in carrying out the analysis. The correlation matrix can also provide a preview of the factor analysis results by identifying separate groups of highly correlated variables.

The extraction phase of factor analysis requires several decisions by the analyst. First a method of factor extraction (Principal components, principal axis factoring, maximum likelihood) must be selected. Principal component analysis been applied for the data considered since it seeks linear combination of variable such that maximum variance is extracted from the variables. Second the number of factors to be retained must be decided. The most frequently used criteria is to keep all factors with eigen values which are greater than one The factor solution is then rotated to make the factors more interpretable. Choices include varimax, quartimax, equimax, and oblique rotation but varimax being the most frequently used technique.

III. DATA ANALYSIS

Data for this study were drawn from the various industrial production from 1997 to 2006 been considered. The data set include Food, Edible Oil, Textile, leather product, Basic Industrial Organic Chemicals, Cosmetics and Toiletries, Tyres and Tubes etc., used for analysis. The Principal Component Analysis was carried out for Cosmetics and Toiletries data set which has 6 variables with 648 observations for illustrative purpose. Pearson product moment correlations between all trait variables were used. As this method provide a unique solution original data can be reconstructed from the results. The solution was obtained by rotating all factors with eigen values greater than one following the rule of Kaiser (1960) and Guttman (1954).

Table-1. Correlation Matrix

	SOAPSSI	DETGENT	HAIROIL	TOOTHPASTE	TOOTHPOWDER	AGARBATH
Correlation	1.000	.558	.427	-.138	-.816	.773
	DETERGENT	1.000	.699	.097	-.366	.563
	HAIROIL	.427	.699	1.000	.151	-.264
	TOOTHPASTE	-.138	.097	.151	1.000	.128
	TOOTHPOWDER	-.816	-.366	-.264	.128	1.000
	AGARBATHI	.773	.563	.128	-.490	1.000

Table [1] presents the analytical process on matrix of correlation between variables for the factors where appropriate variables should be correlated. Principal component analysis requires that there be some correlations greater than 0.30 between the variables included in the analysis. The correlations greater than 0.30 are highlighted in dark in absolute values.

Table-2. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.667
Bartlett's Test of Sphericity	Approx. Chi-Square	360.162
	df	15
	Sig.	.000

Table [2] shows that the Kaiser-Meyer-Olkin measure of sampling adequacy which has been interpreted as .667 that exceeds the minimum requirement 0.50 for overall MSA and Bartlett test of sphericity was significant ($\chi^2_{360.162, DF=15, p<.001}$). Both of these measures indicated that the data were appropriate for factor analysis.

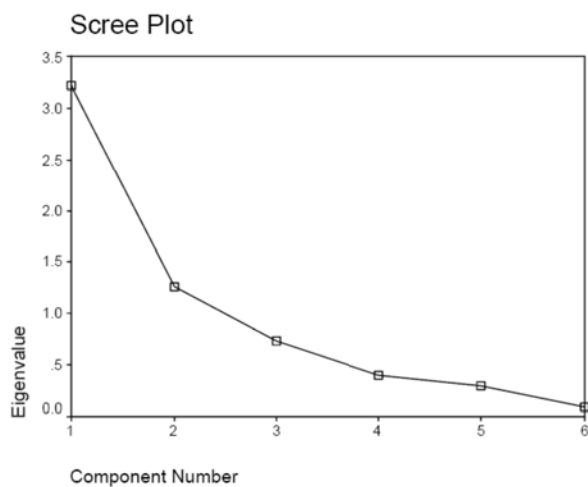


Fig .1 Scree Plot

Scree plot is a graphical plot of the eigen values against the component number. Scree plots are useful for finding an upper bound (maximum) for the number of components that should be retained. Fig .1 is performed by plotting the

eigen values with their respective component number and retaining number of component before break point in the plot. Hence two components been extracted since its eigen values are greater than one.

Table-3. Communalities

	Initial	Extraction
SOAPSSI	1.000	.893
DETERGENT	1.000	.750
HAIROIL	1.000	.749
TOOTHPASTE	1.000	.664
TOOTHPOWDER	1.000	.689
AGARBATHI	1.000	.731

Table [3] shows that the table of communalities represent the proportion of the variance in the original variable is accounted by the factor solution. It shows communalities before and after extraction. Principal component analysis works on the initial assumption that all variance is common; therefore before extraction the communalities are all one. The communalities in the column labeled extraction reflect the common variance in the data structure. It is observed that that 89.3% of the variance associated with SOAPSSI is common or shared variance. After extraction some of the factor are discarded and so some information is lost. The amount of variance in each variable that can be explained by the retained factors is represented by the communalities after extraction. Once any variables with communalities less than 0.50 have been removed from the analysis and the pattern of factor loadings should be examined to identify variable that have complex structure.

Table-4. Component Matrix

	Component	Component
	1	2
SOAPSSI	.907	-.266
DETERGENT	.787	.362
HAIROIL	.709	.496
TOOTHPASTE	.0058	.813
TOOTHPOWDER	-.739	.378
AGARBATHI	.852	.0079

Extraction Method: Principal Component Analysis.

2 components extracted.

The first two principal components are given by

$$Y_1 = .907\text{SOAPSSI} + .787\text{DETERGENT} + .709\text{HAIROIL} + .0058\text{TOOTHPASTE} - .739\text{TOOTHPOWDER} + .852\text{AGARBATHI}.$$

$$Y_2 = -.266\text{SOAPSSI} + .362\text{DETERGENT} + .496\text{HAIROIL} + .813\text{TOOTHPASTE} + .378\text{TOOTHPOWDER} + .00795\text{AGARBATHI}.$$

Principal component analysis seeks to explain the correlation structure of set of predictor variable using smaller set of linear combination . The linear combinations are called components. The total variability of a data set produced by the complete set of m variables can often be accounted for the primarily by a smaller set of k linear combination of these variables which would mean that there is almost as much information in the k component as there are in the original m variables. Principal Component analysis was carried out on the six predictors for the same. The component matrix is shown in table [4] where each column represents one of the components $Y_i = e_i Z$. The cell entries called the components weights represent the partial correlation between the variable and component. Although the component weights do not exactly equal to those of training set, the same 2 components were extracted with one to one correspondence in terms of variable being associated with component. This may be considered for validation of the Principal component analysis performed.

Table-5. Rotated Component Matrix.

	Component	Component
	1	2
SOAPSSI	.944	2.727E-02
DETERGENT	.636	.202
HAIROIL	.521	.398
TOOTHPASTE	-.308	.755
TOOTHPOWDER	.820	.131
AGARBATHI	.834	.188

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.
 Rotation converged in 3 iteration

Table [5] shows the Rotated Component Matrix which is a matrix of the factor loadings for each variable onto each factor. There are several things to consider about the format of this matrix. Complex structure occurs when one variable has high loadings or correlation (.50 or greater) on more than one component. If a variable has complex structure it should be removed from the analysis. Variables are checked for complex structure if there is more than one component in the solution. Variables that load on only one component are described as having simple structure. In table [5] none of the variables demonstrated complex structure. Hence, it is not necessary to remove any additional variable because of complex structure.

Table-6. Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.219	53.643	53.643	3.219	53.643	53.643	3.031	50.513	50.513
2	1.257	20.958	74.601	1.257	20.958	74.601	1.445	24.088	74.601
3	.740	12.327	86.928						
4	.402	6.707	93.635						
5	.292	4.874	98.510						
6	8.943E-02	1.490	100.000						

Extraction Method: Principal Component Analysis.

Table [6] shows the eigen values associated with each factor represent the variance by that particular linear component and also displays the eigen values in terms of the percentage of variance explained. In the last part of the table, the eigen values of the factors after rotation are shown. Rotation has the effect of optimizing the factor structure and for the data set considered, two components were extracted which explain 74.601% of the total variance in the variables which are included on the components.

Now the data is being split up into two parts in the ratio of 3:1. Three fourth of data set is being utilized for modeling and remaining one fourth is being used for validating the same.

Table-7. Communalities

	Initial	Extraction
SOAPSSI	1.000	.887
DETERGENT	1.000	.754
HAIROIL	1.000	.727
TOOTHPASTE	1.000	.681
TOOTHPOWDER	1.000	.664
AGARBATHI	1.000	.664

Extraction Method: Principal Component Analysis.

Only cases for which SPLIT = 0 are used in the analysis phase.

Table-8. Communalities

	Initial	Extraction
SOAPSSI	1.000	.909
DETERGENT	1.000	.735
HAIROIL	1.000	.864
TOOTHPASTE	1.000	.725
TOOTHPOWDER	1.000	.710
AGARBATHI	1.000	.783

Extraction Method: Principal Component Analysis.

Only cases for which SPLIT = 1 are used in the analysis phase.

The communalities for both split samples displayed in tables [7] and [8] satisfy minimum requirement being larger than 0.50 and hence satisfied the validation.

Table-9. Rotated Component Matrix

	Component	Component
	1	2
SOAPSSI	.640	-.391
DETERGENT	.865	-7.477E-02
HAIROIL	.150	.850
TOOTHPASTE	.307	.766
TOOTHPOWDER	-.522	.397
AGARBATHI	.677	-.453

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

- Rotation converged in 3 iterations.
- Only cases for which SPLIT = 0 are used in the analysis phase.

Table-10. Rotated Component Matrix

	Component	Component
	1	2
SOAPSSI	.952	5.627E-02
DETERGENT	.617	.295
HAIROIL	.472	.801
TOOTHPASTE	-.303	.796
TOOTHPOWDER	-.838	8.917E-02
AGARBATHI	.858	.218

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

- Rotation converged in 3 iterations.
- Only cases for which SPLIT = 1 are used in the analysis phase.

Both Tables [9] and [10] shows the pattern of factor

loading which shows the variables SOAPSSI; DETERGENT; and TOOTHPOWDER; AGARBATHI loading on the first component and HAIROIL; TOOTHPASTE loading on the second component which met sufficient criteria for simple structure.

IV. INTERPRETING THE VALIDATION RESULTS

The communalities in both split sample validation met the criteria. The pattern of loadings for the both validation sample is the same as the pattern for the analysis using the full data set. This validation analysis supports a finding that the results of this rotated component analysis are generalizable. Also to check outliers which are having an impact on the factor solution factor score been computed. If any values in the factor score that are larger than +/- 3.0 in absolute values reveals an outlier. While computing factor score it has been observed that an outlier lies in the year 1999 in the month of April since it exceeds the same.

V. CONCLUSION

It is important to develop a model for each of the set to carry out prediction. Identification of outlier will help in developing suitable models either by removing the outlier or retaining the outliers. The developed model will be more suitable for decision makers in evolving suitable guidelines or policy changes in the light of identifying data.

REFERENCES

- [1] J.Stevens, (1996), Applied Multivariate Statistics for the Social Sciences. Lawrence Erlbaum Associates New Jersey USA.
- [2] Lawlay D.N. & Maxwell A.E.(1963) Factor Analysis as a Statistical Method London Butterworth.
- [3] Solomon H (1960) " A Survey of Mathematical Models in factor analysis". Mathematical Thinking in the Measurement of Behaviour. New York: Free Press.
- [4] Hair J. Anderson R Tatham R. and Black W (1998) Multivariate Data Analysis. Prentice Hall. New Jersey USA.
- [5] Richard A. Johnson and Dean Wichern (1998) Applied Multivariate Statistical Analysis, Prentice Hall, Upper Saddle River.



Mr. Sivakumar - Research Scholar, Sathyabama University. He has 10 years of academic experience and has presented 9 papers in national and international conferences.