9

FACTOR ANALYSIS: A TECHNIQUE FOR DATA REDUCTION

P.S. Swathi Lekshmi

Introduction:

The basic purpose of factor analysis is to summarize data so that relationship and patterns can be easily interpreted and understood. It is normally used to regroup variables into a limited set of clusters based on shared variance. Hence, it helps to isolate constructs and concepts. Factor analysis uses mathematical procedures for the simplification of interrelated measures to discover patterns in a set of variables (Child, 2006). Attempting to discover the simplest method of interpretation of observed data is known as parsimony, and this is essentially the aim of factor analysis (Harman, 1976).

Factor analysis has its origins in the early 1900's with Charles Spearmen's interest in human ability and his development of the Two-Factor theory; this eventually lead to a burgeoning of work on the theories and mathematical principles of factor analysis (Harman,1976). Factor analysis is used in many fields such as behavioural and social sciences, medicine, economics, and geography as a result of the technological advancements of computers.

Uses of Factor Analysis:

Factor analysis is useful for studies that involve a few or hundreds of variables, item from questionnaires or a battery of tests which can be reduced to a smaller set, to get at an underlying concept, and to facilitate interpretations. It is easier to focus on some key factor rather than having to consider too many variables that may be trivial, and so factor analysis is useful for placing variables into meaningful categories. Many other uses of factor analysis include data transformation, hypothesis-testing, mapping, and scaling (Rummel, 1970).

This technique is applicable when there is a systematic interdependence among a set of observed or manifest variable and the researcher is interested in finding out something more fundamental or latent which creates communality (commonness).

The recommended sample size is at least 300 participants and the variables that are subjected to factor analysis each should have at least 5 to 10 observations (Comery and Lee, 1992).

Factor Analysis- Methodology Framework

The theoretical basis for factor analysis is that variables are correlated because they share one or more common components. That is correlations among variables are explained by underlying factors. Mathematically a one-factor model for three variables can be represented as follows (Vs are variables Fs are factors Es represent random error).

$$V_1 = L_1 * F_1 + E_1$$
$$V_2 = L_2 * F_1 + E_2$$
$$V_3 = L_3 * F_1 + E_3$$

Each variable is composed of the common factor (F_1) multiplied by a loading coefficient ($L_1L_2L_3$-the lambdas) plus a random component. If the factor were directly measurable (which it isn't) this would amount to a simple regression equation. Since these equations cannot be solved as given (the Ls, Fs and Es are unknowns), factor analysis takes an indirect approach. If the equations above hold, then consider why variables $V_1$and V_2 correlated. Each contains an error (the Es are assumed to be random or unique) component that cannot contribute to their correlation (errors are assumed to have O correlation). However they share the factor F_1 so if they correlate their correlation should be related to L_1 and L_2 (the factor loadings). If this logic is applied to all the pairwise correlations, the loading coefficients can be estimated from the correlation data. Thus one factor might account for the correlations in a set of variables. If not, the equations can be easily generalized to accommodate additional factor. There are different approaches to fitting factors to a correlation matrix (least squares, generalized least squares, maximum likelihood, etc.) which have given rise to a number of factor methods. A basic assumption of factor analysis is that the variables used in factor analysis are linear combinations of some underlying factors.

The idea of a principal component

A concept related to most methods of factoring is the idea of a principal component. A principal component is a linear combination of observed variables that is independent (orthogonal) of other components. The first principal component accounts for the largest amount of variance in the input data. The second component accounts for the largest amount of the remaining variance in the data and so on.

Varimax rotation

The ideal result of rotation is that each variable will have a high loading on a single factor (have a lambda coefficient near one) and small loading (near zero) on the other factors. Therefore, the net effect of rotation as well as its main motivation is to facilitate interpretation.

Varimax rotation attempts to simplify interpretation by maximizing the variances of the variables loadings on each factor (i.e., tries to simplify the factors).

Application of Factor Analysis in Fisheries sector: An example

In the present study, 15 profile characteristics of shrimp farmers in Nellore district of Andhra Pradesh, and one dependent variable namely the extent of adoption of shrimp culture technologies were used.

Factor loadings of profile characteristics with respect to extent of adoption of shrimp culture technologies

The results from the factor analysis explained the number and nature of relationship existing among the profile characteristics with the extent of adoption of shrimp culture technologies and the results are presented in Table 1.

Table 1. Factor loadings of profile characteristics with respect to extent of adoption of shrimp culture technologies (n = 60)

Sl. No.	Profile characteristics	Factor I	Factor II	Factor III	Factor IV	Communality
1.	Age	0.475	0.023	-0.1123	0.763	0.821
2.	Education	0.820	0.107	-0.092	-0.291	0.778
3.	Occupation	0.859	0.193	-0.087	-0.174	0.812
4.	Farm size	0.757	-0.069	0.022	0.371	0.7016
5.	Experience in shrimp farming	0.541	-0.373	0.583	0.066	0.776
6.	Annual income	0.049	0.795	0.279	0.179	0.744
7.	Family size	-0.090	0.787	0.394	0.118	0.797
8.	Ownership of shrimp farm	0.813	-0.250	0.058	0.181	0.760

9.	Marketing behavior	0.850	-0.151	0.244	-0.127	0.821
10.	Material possession	0.635	-0.027	0.515	-0.248	0.730
11.	Social participation	0.883	0.051	-0.167	-0.198	0.850
12.	Information seeking behavior	0.743	0.261	-0.231	-0.188	0.709
13.	Extension participation	0.506	0.419	-0.298	-0.065	0.525
14.	Economic motivation	0.557	0.009	-0.496	0.141	0.577
15.	Risk orientation	0.506	-0.169	0.118	0.194	0.336
Eigen values		6.454	1.806	1.360	1.132	
% of variation explained		43.029	12.040	9.069	7.548	
Cumulative % variation explained		43.028	55.068	64.137	71.685	

A close perusal of Table 1 gives the factor loadings, communalities, eigen values, and the percentage of variance explained by the factors. It could be seen from the table, that out of the 15 profile characteristics, five factors have been extracted and these five factors, together explain the total variance of these profile characteristics to the extent of 71.68 per cent.

The factors extracted as such are rarely interpretable and have only theoretical significance. It is therefore, necessary to rotate the factors, so that the rotated factors may be meaningfully interpreted. The varimax rotation was used to obtain meaningful interpretation, and the results are given in Table 2.

Table 2: Rotated factor (varimax) matrix of fifteen profile characteristics

Sl. No.	Profile characteristics	Factors			
		1	2	3	4
1.	Age	0.123	0.080	**0.889**	0.095
2.	Education	**0.779**	0.408	0.061	-0.028
3.	Occupation	**0.791**	0.391	0.173	0.070
4.	Farm size	0.376	0.422	**0.629**	0.010
5.	Experience in shrimp farming	-0.040	**0.853**	0.200	-0.082
6.	Annual income	0.108	-0.040	0.062	**0.852**
7.	Family size	-0.027	-0.024	-0.066	**0.890**
8.	Ownership of shrimp farm	0.407	0.565	0.489	-0.191
9.	Marketing behavior	0.492	0.734	-0.185	-0.081
10.	Material possession	0.293	0.789	-0.065	0.132
11.	Social participation	**0.807**	0.393	0.187	-0.096
12.	Information seeking behavior	**0.805**	0.198	0.134	0.068
13.	Extension participation	**0.675**	-0.055	0.151	0.209
14.	Economic motivation	**0.590**	-0.071	0.428	-0.201
15.	Risk orientation	-0.179	0.405	0.365	-0.080
Eigen values		4.030	3.080	1.938	1.704
% of variation explained		26.869	20.535	12.919	11.363
Cumulative % variation explained		26.869	47.404	60.322	71.685

An analysis of Table 2 shows the interpretation of the rotated factors in the varimax matrix. A total of four factors have been identified as having maximum percentage variance. Each factor column was scanned for identifying a few profile characteristics with significant high loadings. Thus from each factor column, the profile characteristics having a factor loading of more than **0.5** were selected. Thus the selected factor loadings from each factor column was selected and presented in Table 3.

Table 3: Profile characteristics with factor loadings under different factors for extent of adoption of shrimp culture technologies

Factor	Profile characteristics	Factor loadings
FACTOR I	Education	0.779
	Occupation	0.791
	Social participation	0.807
	Information Seeking behaviour	0.805
	Extension participation	0.675
	Economic motivation	0.590
FACTOR II	Experience in Shrimp farming	0.853
	Ownership of Shrimp farm	0.565
	Marketing behavior	0.734
	Material possession	0.789
FACTOR III	Age	0.889
	Farm Size	0.629
FACTOR IV	Annual income	0.852
	Family size	0.890

An analysis of Table 3 shows the groupings of the profile characteristics under each factor with respect to their factor loadings.

FACTOR I

The profile characteristics in the factors were identified as prime factor which explained 43.03 per cent of variance on the overall extent of adoption of technologies by shrimp farmers. These include social participation (0.807), information seeking behaviour (0.805), occupation (0.791) education (0.779), extension participation (0.675) and economic motivation (0.590). It could be seen from the table that the profile characteristics, social participation and information seeking behaviour had highest factor loadings followed by Education. Hence, this factor is labeled as "*socio-personal*" factor.

FACTOR II

From Table 3, it could be further noted that there were 4 characteristics which had significant loadings on factor III. They were experience in shrimp farming (0.853), material possession (0.789), marketing behaviour (0.734) and ownership of shrimp farm (0.565).All these characteristics are of personal importance and hence it has been labeled as "personal" factor. The second factor accounted for 12.04 per cent of the total variance.

FACTOR III

Age and farm size under this factor accounted for 9.07 per cent of the total variance. Of these two, age had a higher factor loading of 0.889, and hence this factor was termed as "individual" factor.

FACTOR IV

The two profile characteristics which had significant loadings on factor IV were family size and annual income. This factor accounted for 7.55 per cent of the total variance; and hence this factor was termed as “family” factor.

Conclusion:

In this study, factors analysis was used to group the variables into factors based on the communalities observed, and to find out the relative importance of each factor in accounting for the particular set of variables being analysed. The method of factor analysis used for the study was principal component analysis and the rotation method was varimax rotation. It could be inferred from the foregoing study that the *socio-personal* factor accounted for the maximum percentage of the total variation on the overall extent of adoption of technologies by shrimp farmers.

Suggested Readings:

- Child, D. (2006). The essentials of factor analysis. (3rd ed.). New York.
- Comrey, L.A. & Lee, H.B. (1992). *A first course in factor analysis* (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates.
- Harman, H.H. (1976). *Modern factor analysis (3rd ed. revised)*. Chicago, IL: University of Chicago Press, NY: Continuum International Publishing group.
- Rummel, R. J. (1970). *Applied factor analysis.* Evanston, IL: Northwestern University Press.
- Swathi Lekshmi, P. S. and Chandrakandan, K. (2005) Personality factors influencing the adoption of shrimp culture technologies. Journal of Extension Education, 16 (3 & 4). pp. 3802-3806.