

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

A Comprehensive Review of Factor Analysis

Hyeuk Kim¹

Assistant Professor, Division of Global Management Engineering, Hoseo University, Asan, Korea¹

ABSTRACT:Factor analysis is one of the dimension reduction methods. It is widely used in any areas even though its history is very long. Researchers and engineers are applying factor analysis much as they focus on massive data and start to analyse. We review factor analysis in the paper. All aspects of factor analysis are studied from the theoretical part to an experimental part. Especially, a factor rotation method affects the result of factor analysis and we verify it from the experiment.

KEYWORDS:Factor analysis, Principal component analysis, Factor extraction method, Factor rotation method

I. INTRODUCTION

Factor analysis discovers the interrelationship among the corresponding variables through the covariance and the correlation [1]. The correlation and the structure of the variables and the questions are revealed through the method and they are combined into a few factors based on the information of the corresponding variables. Many variables are compressed into a small number of intrinsic factors through factor analysis. The basic concept of factor analysis is to minimize loss of information by reducing the number of factors from the original number of its variables. The advantages of factor analysis are that it helps a researcher understand information and process an extra experiment. On the other hand, the disadvantage of factor analysis is that it is hard to understand the factors' meanings if the derived factors are generated randomly. The result must be interpreted cautiously. It can lead to misinterpretation if we classify and interpret the combined groups of variables without theoretical background of validity and reliability. The paper consists of 4 parts. We introduce factor analysis in short in the first section. In the second section, we investigate several aspects of factor analysis. We make an experiment for factor analysis in the third section. In the last section, we make a conclusion.

II. VARIOUS ASPECTS OF FACTOR ANALYSIS

The characteristics of factor analysis are as follows. There is no distinction between a dependent variable and independent variables. The general statistical analyses consider the population estimation in the last procedure, but the factor analysis does not consider inference procedure. After the variables are reduced to a few factors, factors are available for such analyses as regression, clustering, correlation, and so on. There are two types of factor analysis. R-type factor analysis is to create a few factors based on the variables and it is a widely used method. Q-type factor analysis is to classify individuals into several reciprocal groups. Cluster analysis is generally applied to Q-type factor analysis since it is hard to calculate in factor analysis approach. We only investigate R-type factor analysis in the research. Meanwhile, we classify factor analysis in a different view. There are another two styles of factor analysis. An exploratory factor analysis and a confirmatory factor analysis. In the research, we only consider an exploratory factor analysis. The disadvantages of an exploratory factor analysis are as follows. The number of factors and the factor rotation method are decided by a researcher's subjectivity even though several subsidiary tools are offered for decision. The variables in the same factors usually have similar characteristics. However, the variables in the factors generated from exploratory factor analysis sometimes is hard to find common concept and interpret the factor. Therefore, the interpretation of factors is in danger because a researcher decides subjectively.

Factor analysis cannot be applied in any condition. Factor analysis analyse the variables of which forms are interval or ratio scales. An appropriate data for factor analysis is basically continuous data. In survey research, Likert scale is used. Data from Likert scale is regarded as continuous data since the value from Likert scale belongs to ordinal number. The

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

minimum number of the samples is 50 and the recommended number of the samples is over 100. Another recommendation is that the number of the samples is over two times of the number of variables. As the number of the samples increases, the result of factor analysis becomes better like other statistical analyses.

Table 1. Process for factor analysis

Step	1	2	3	4
Topic	Topic decision	Factor extraction method	Factor rotation method	Calculation of the factor score and interpretation of the result
Subtopics	<ul style="list-style-type: none"> - Decide research topic - Decide the sample size - Develop the questions in survey - Decide the number of factors roughly from the previous research - Investigate/collect/code data 	<ul style="list-style-type: none"> -Decide the factor extraction method -Decide the number of factors -Examine the result without factor rotation 	<ul style="list-style-type: none"> -Decide the factor rotation method -Examine the result with factor rotation 	<ul style="list-style-type: none"> -Test the model through KMO and Bartlett's test -Interpret communalities and total variance explained -Interpret the component matrix -Decide the final factors and name them -Generate the factor scores

Table 1 describes the procedure for applying the factor analysis to data. The main process in the second stage is to decide the factor extraction method in factor analysis. There are several factor extraction methods. A principal component analysis [2] forms uncorrelated linear combinations from the variables. The order of the components is decided by the size of their variances. The first component has the largest variance. The second component has the second largest variance [3]. It is a common factor extraction method. An unweighted least-squares method minimizes the total sum of the squared differences between the original and the modified correlation matrices [4]. A generalized least-squares method is different from an unweighted least-squares method. It considers the weights in correlations. The variables with high uniqueness have little weights and the variables with low uniqueness have high weights. A maximum-likelihood method estimates parameters by considering the corresponding maximum likelihood [5].

It is one of the most important task to decide the number of the factors in factor analysis. We include the factor into the final form until the corresponding eigenvalue is equal to or greater than 1. An eigenvalue means the sum of the squared factor loading values. It is the power of exploration which is same with the variation of the variance. Another method is to use Scree plot. Scree plot is the plot with the number of factors as x-axis and the eigenvalues of the corresponding factors as y-axis. The eigenvalue decreases as the number of factors increases. The number of factors is decided at the point which the line is folded suddenly [6][7].

In the third stage in Table 1, we rotate the axes of factors for better performance. It effects the performance much. The basic distinct is orthogonal rotation or oblique rotation. The basic hypothesis for two rotations are completely different. A research must select the specific method carefully. The factor rotation methods are described in Table 2 [8].

Table 2. Methods for factor rotation

Rotation type	Method	Description
Orthogonal rotation	Varimax [9]	Factors are independent in rotating. There is no correlation between factors. It is easy to interpret the result since the result is simple.
	Quartimax [10]	
	Equimax	
Oblique rotation	Oblimin	Factors are related each other. The value of the corresponding correlation is not zero. It is hard to interpret the result since the form of result is complex.
	Promax [11]	

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

There is a reason to apply factor rotation before factor analysis. There is a limitation to verify the levels between variables with factor loading values. Therefore, we rotate the directions of the factors based on the mathematical rule first. It makes the factor with a high factor loading value gets greater factor loading value up to 1 and the factor with a low factor loading value gets lower factor loading value up to 0 [12]. Factor rotation reduces the dimensions of the variables through factor loading values which mean the corresponding correlation coefficients.

III. EXPERIMENTAL RESULTS

We use fictional data to compare the result of factor analysis. It is the survey about a merchandise which consists of 15 questions. 325 people answers the survey. The result with the principal component analysis and no rotation is as follows.

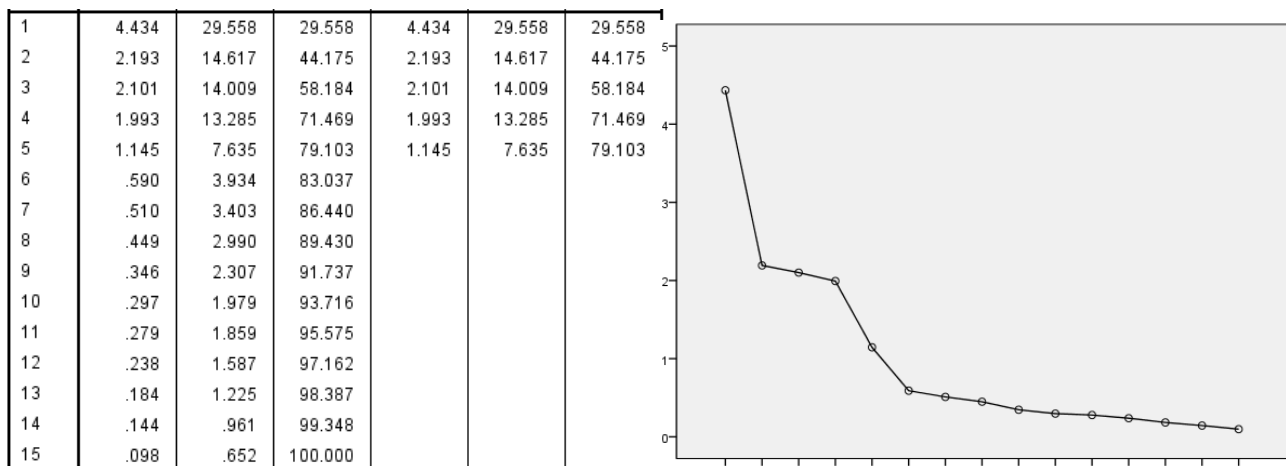


Fig. 1. Total variance explained and Scree plot for factor analysis with principal component analysis and no rotation

The appropriate number of factors is 5 when we decide it from the above table and Scree plot. The component matrix is described in Figure 2. The factor consists of the questions with high values. We can assign questions into the factors, but it is not easy because the values are not different much in some factors. The row means the variable and the column means the factor in Figure 2. Next, we rotate the directions of factors and apply the factor analysis.

	1	2	3	4	5
1번	.634	.053	-.683	-.126	.114
2번	.659	.042	-.678	-.139	.102
3번	.665	.036	-.677	-.118	.061
4번	.302	-.315	.066	.668	.218
5번	.360	-.206	.105	.610	.375
6번	.429	-.157	.120	.623	.246
7번	.502	-.415	.387	-.411	.284
8번	.536	-.328	.375	-.487	.219
9번	.530	-.317	.320	-.500	.142
10번	.443	.722	.300	-.039	.127
11번	.431	.758	.281	.023	.084
12번	.465	.695	.231	.043	.089
13번	.635	-.130	.092	.244	-.435
14번	.651	-.061	.196	.087	-.495
15번	.710	-.201	.153	.098	-.505

Fig. 2. Component matrix for factor analysis with principal component analysis and no rotation

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

Varimax method is the widely-used rotation method. Total variance explained and Scree plot from the factor analysis with principal component analysis and Varimax rotation are described in Figure 3. They are exactly same with Figure 1. They are always same regardless of factor extraction method and factor rotation method. Therefore, we omit the table and the plot like Figure 3.

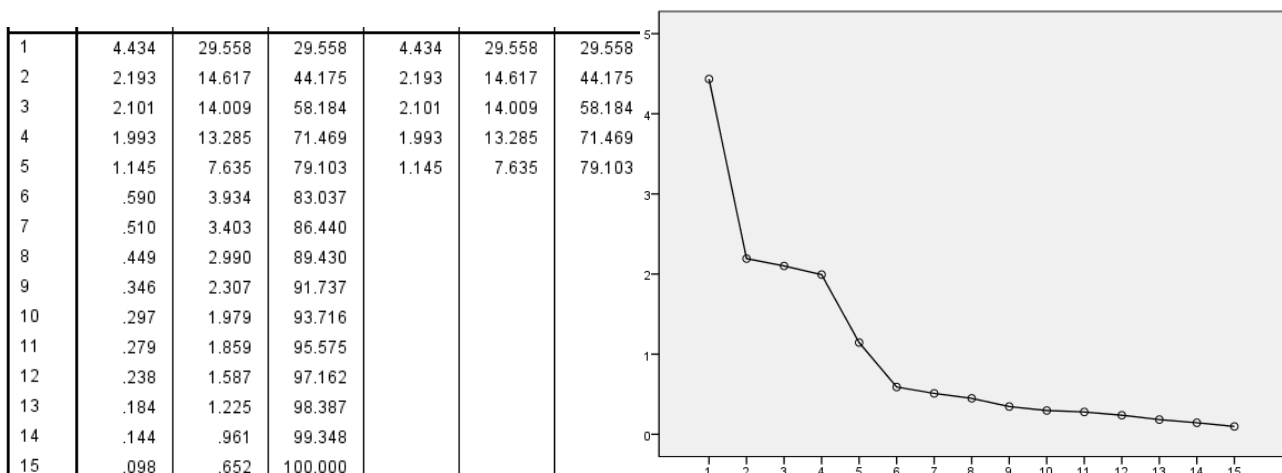


Fig. 3. Total variance explained and Scree plot for factor analysis with principal component analysis and Varimax rotation

Figure 4 describes the component matrices from Varimax rotation and Quartimax rotation. The rotated component matrices are dramatically different from the original component matrix. However, they are almost same each other.

	1	2	3	4	5		1	2	3	4	5
1번	.936	.077	.070	.106	.055	1번	.936	.077	.070	.104	.055
2번	.945	.076	.091	.129	.051	2번	.945	.076	.091	.127	.051
3번	.937	.068	.070	.169	.052	3번	.938	.068	.071	.167	.053
4번	.019	-.095	-.018	.161	.808	4번	.019	-.094	-.018	.160	.809
5번	.056	.057	.066	.048	.826	5번	.056	.057	.066	.046	.826
6번	.062	.109	.030	.185	.788	6번	.063	.109	.031	.184	.788
7번	.039	.001	.893	.104	.117	7번	.039	.001	.894	.102	.117
8번	.073	.073	.888	.145	.013	8번	.074	.073	.889	.143	.013
9번	.107	.050	.838	.191	-.043	9번	.107	.050	.838	.189	-.043
10번	.054	.899	.090	.073	.003	10번	.054	.899	.090	.071	.003
11번	.049	.913	.011	.102	.016	11번	.049	.913	.011	.101	.016
12번	.102	.857	.021	.121	.056	12번	.102	.857	.021	.119	.056
13번	.157	.087	.080	.764	.236	13번	.158	.088	.082	.763	.238
14번	.099	.175	.174	.800	.084	14번	.101	.176	.175	.799	.085
15번	.156	.061	.229	.856	.139	15번	.158	.063	.231	.855	.140

Fig. 4. Component matrix for factor analysis with principal component analysis (a) Varimax rotation method (b) Quartimax rotation method

The results from oblique rotation approach are shown in Figure 5. They are the component matrices from Oblimin rotation and Promax rotation.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

	1	2	3	4	5		1	2	3	4	5
1번	-.001	.007	-.957	.007	.031	1번	.957	.007	-.001	-.033	.007
2번	.016	.002	-.962	-.002	.008	2번	.962	.002	.016	-.010	-.001
3번	-.012	-.010	-.949	-.007	-.043	3번	.950	-.010	-.013	.042	-.007
4번	-.043	-.117	.021	.810	-.073	4번	-.024	-.118	-.045	.065	.813
5번	.060	.047	-.021	.847	.091	5번	.018	.047	.060	-.102	.851
6번	-.004	.085	-.008	.783	-.070	6번	.004	.084	-.006	.062	.786
7번	.916	-.035	.026	.103	.042	7번	-.030	-.037	.920	-.051	.103
8번	.899	.032	-.003	-.013	-.009	8번	.000	.030	.903	.002	-.015
9번	.837	.002	-.037	-.080	-.074	9번	.034	.000	.839	.069	-.082
10번	.055	.909	.011	-.013	.026	10번	-.013	.910	.055	-.029	-.014
11번	-.032	.923	.018	-.004	-.017	11번	-.020	.924	-.034	.015	-.005
12번	-.027	.859	-.037	.032	-.029	12번	.035	.860	-.029	.027	.031
13번	-.062	-.009	-.039	.106	-.792	13번	.030	-.012	-.073	.800	.102
14번	.030	.078	.033	-.057	-.841	14번	-.041	.076	.019	.850	-.062
15번	.078	-.049	-.021	-.011	-.891	15번	.012	-.052	.067	.899	-.017

Fig. 5. Component matrix for factor analysis with principal component analysis (a) Oblimin rotation method (b) Promax rotation method

IV. CONCLUSION

Factor analysis is very popular method nowadays. It reduces the number of features and the approach is especially useful when a researcher handles huge data. Even though they belong to factor analysis category, there are many different factor analysis methods. Therefore, it is very important for a researcher to make several decisions for better performance of factor analysis. Their decisions are the number of factors, the factor extraction method, the factor rotation method, the interpretation of factors, and so on. We discover the differences from the selection of the rotation methods in the paper. We can perform factor analysis in a right manner when we follow the process for factor analysis which is described Table 1.

REFERENCES

- [1] J.-O. Kim, and C. W. Mueller, "Factor analysis: Statistical methods and practical issues", SAGE Publications, 1978.
- [2] K. Pearson, "On lines and planes of closest fit to systems of points in space", Philosophical Magazine, 2(11), pp. 559-572, 1901.
- [3] G. H. Duntzman, "Principal components analysis", SAGE Publications, 1989.
- [4] P. Paatero, "Least squares formulation of robust nonnegative factor analysis", Chemometrics and Intelligent Laboratory Systems, 37(1), pp. 23-35, 1997.
- [5] C.-Y. Liou, and B. R. Musicus, "Cross entropy approximation of structured gaussian covariance matrices", IEEE Transactions on Signal Processing, 56(7), pp. 3362-3367, 2008.
- [6] R. D. Ledesma, and P. Valero-Mora, "Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis", Practical Assessment Research & Evaluation, 12(2), pp. 1-11, 1999.
- [7] W. F. Velicer, "Determining the number of components from the matrix of partial correlations", Psychometrika, 41, pp. 321-327, 1976.
- [8] M. A. Pett, N. R. Lackey, and J. Sullivan, "Making sense of factor analysis: The use of factor analysis for instrument development in health care research", SAGE Publications, 2003.
- [9] H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis", Psychometrika, 23(3), pp. 187-200, 1958.
- [10] J. E. Jackson, "Quartimax rotation", Encyclopedia of Biostatistics, 2005.
- [11] A. E. Hendrickson, and P. O. White, "Promax: A quick method for rotation to oblique simple structure", British Journal of Mathematical and Statistical Psychology, 17(1), pp. 65-70, 1964.
- [12] P. Kline, "An easy guide to factor analysis", Routledge, 1993.