

## Research Article

### Improved Principal Component Analysis and its Application in the Evaluation of the Industrial Structure

LI Fen-Hong

Department of Mathematics and Computational Science, Shangluo College, Shangluo, 726000, China

**Abstract:** In this study, improved principal component analysis method is put forward to avoid the shortage of comprehensive evaluation dealt with principal component analysis method. When the contribution rate of the first principal components is short of asks, we can choose to rotate factor loading matrix and select multiple main components and synthesize and weigh the variation coefficient and variance contribution as weight coefficient, to set up a comprehensive evaluation model. As an example of the main indicators of Shangluo city Industrial data, the comparative study uses factor analysis and improved model, systematic classification with cluster analysis. The result shows that comprehensive evaluation of the improved model in the practical problems is more reasonable and objective.

**Keywords:** Clustering analysis, factor analysis, industrial structure principal component analysis

#### INTRODUCTION

Keep the introduction short. Precise the introduction in a significant manner and exclude the all subheadings. The basic principles of research, background earlier work and the purpose of the present studies should be described in the introduction. Introduction should be briefly justifies the research, specifies the hypotheses to be tested and gives the objective (s). Extensive discussion of relevant literature should be included in the discussion.

The principal component analysis is put forward as multiple data dimension reduction processing technology in the early and is widely used in the natural, biology, medicine, management, economy, society, etc. But it exists problem in the actual application. Mainly includes: only take the first principal component or take multiple main composition weighted when using principal component analysis in comprehensive evaluation (Ye, 2006); the characteristics and deficiency of the traditional linear principal components (Gao and Cai, 2004); integrated rank high related indicators appear using principal component analysis (Ye, 2001), improved main component method of comprehensive evaluation (Xu and Wang, 2006; Yan, 1998; Sun and Qian, 2009); the comparison of main component of the comprehensive evaluation method and factor analysis of comprehensive evaluation (Wang and Chen, 2006; Shi, 2007), etc. We all have certain understanding for these problems. The main problems focus on the contribution of first principal components did not reach the required, only choose the first principal component to sort or choose multiple main ingredients to variance

contribution will be multiple main composition weighted before the evaluation; Another is to the improvement of the principal component analysis, combine factor analysis and the clustering analysis method and trying to find some new weighted function instead of variance contribution ratio (Li and Liu, 2010).

Based on the above research and analysis, this study synthesis the variation coefficient and variance contribution as a synthetic weight coefficient, combined with the factor analysis, thus to set up a comprehensive evaluation model and sort and classify with principal component clustering method. As an example of the main indicators of Shangluo city Industrial data, the comparative study uses factor analysis and improved model and the clustering analysis to classify system.

#### METHODOLOGY

**The feasibility of coefficient of variation as the weight coefficient:** Coefficient of variation and multiple correlation coefficient are commonly used as weights in the comprehensive evaluation (Hu and He, 2000), Coefficient of variation reflects the degree of variation between the variables. The original data usually includes two aspects of information: variability information of the degree of variation of each index and the information of the mutual influence degree between the various indicators, which is the coefficient of variation and correlation coefficient of each index. Covariance matrix fully characterizes all of the information of the original data. Its diagonal element is the variance of each indicator and non-diagonal

Table 1: Total variance explained

Component	Initial Eigen values			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	7.522	83.581	83.581	7.522	83.581	83.581	4.743	52.696	52.696
2	0.700	7.778	91.359	0.700	7.778	91.359	3.480	38.663	91.359

elements contain the information of the correlation coefficient between the various indicators.

The original data are often immeasurable steel in a comprehensive evaluation, such as the mean and the proportion method. Diagonal elements of data covariance matrix that is immeasurable steel relate to coefficient of variation of the variable. In the principal component analysis, the main diagonal elements of the covariance matrix that is the standardized data usually are all 1, which in fact denies the degree of variation of the index variation information. Standardizing the raw data, the principal component that comes from the covariance matrix or correlation coefficient matrix reflects the correlation of the data. But the coefficient of variation, which is the information of the data itself and can highlight the relative changes of every index range, should be reflected in the evaluation results. The comprehensive evaluation shows the relationship and difference information of the data, which makes all information of the original data reflected in comprehensive evaluation and so as to achieve the purpose of comprehensive evaluation to the problem.

**Improved principal component analysis:** On the basis of the principal component analysis, the principal component analysis is improved as follow:

Standardizing the original data, we get data matrix  $X = (x_{ij})_{n \times p}$ . Anglicizing the data by using the principal component analysis, we are orthogonal rotating if the contribution rate of the first principal component can not reach the requirements and choose principal component if its contribution rate more than 85%. In establishing comprehensive evaluation model, we choose the variable coefficient of variation and variance contribution ratio of the principal component as the two weights and weighting .So we get the comprehensive evaluation model

$$E = \sum_{k=1}^p \frac{\partial_k \rho_k}{\sum_{i=1}^p \partial_i \rho_i} X_k \tag{1}$$

where,  $\partial_k$  is a principal components variance contribution.  $W_k = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i}$  is the coefficient of the variable  $X_k$ ,  $\rho_k$  is the variable coefficient of variation  $X_k$ .

**The comprehensive evaluation of the industrial structure of shangluo city:**

**Establishment of index system and data sources:** Combined with the actual situation in shangluo, we choose  $X_1$ -gross industrial output value,  $X_2$ -industrial

Table 2: Component matrix and rotated component matrix

Variable	Component matrix		Rotated component matrix	
	$f_1$	$f_2$	$f_1$	$f_2$
$X_1$	0.910	0.109	0.631	0.665
$X_2$	0.957	*	0.730	0.619
$X_3$	0.960	-0.163	0.844	0.486
$X_4$	0.954	*	0.725	0.620
$X_5$	0.820	0.498	0.314	0.907
$X_6$	0.889	0.418	0.417	0.889
$X_7$	0.943	-0.277	0.903	0.389
$X_8$	0.857	-0.346	0.881	0.281
$X_9$	0.926	-0.202	0.842	0.436

"\*" indicates that data is greater than or equal to -0.100 is less than or equal to 0.100

sales value,  $X_3$ -assets worth,  $X_4$ -advocate business income,  $X_5$ -total profit,  $X_6$ -profit tax amount,  $X_7$ -this year deal with total wages,  $X_8$ --welfare and  $X_9$ -the number of the average of the total amount of the employees as the research object.

The original data of index come from Shangluo Statistical Yearbook (2009).

**Improved evaluation model of the principal component analysis:**

We get KMO test value 0.734 by factor analysis for the data by SPSS11.5 (Lu, 2007) and using the principal component analysis when we extract the factors, which show that it is suitable for factor analysis. Obtained by the output of each common factor corresponding to the eigenvalues, contribution rate of the common factor, the cumulative contribution ratio, the output results are shown in Table 1. The first principal component's contribution rate is 83.581%, which is not up to the requirements, so using Varlmax method rotates the factor loading matrix and get accumulative contribution rates of the first and second principal components (i.e., 2 integrated indexes) reached 91.359% after rotation. We select the first and second principal components and use the improved model (1) to calculate the comprehensive score of the structure of industry in Shangluo City and rank.

**Factor analysis of process and outcome evaluation:**

According to the analysis, to extract two common factors, the rotated factor loading matrix in Table 2. From Table 2, the rotation in the nine indicators  $X_2, X_3, X_4, X_7, X_8, X_9$  the load on a common factor and these indicators, industrial sales output value, total assets, the main business income this year to cope with the total wages, welfare total, all employees in the annual average number of information, basically reflects the scale of industrial development, so the first factor named for the scale factor ( $f_1$ ); The load on the second factor value of these indicators is the total industrial

Table 3: Factor score and ranking

Ranking	Factor $f_1$	Factor $f_2$	Scores of the $F$	Comprehensive factor $F$	Score of the $E$	Model (1-1) $E$
1	Iron ore mining	Tungsten and niobium mining and dressing	2.1907	Iron ore mining	7.1774	Iron ore mining
2	Chinese patent medicines Manufacture	Other metal ores Mining	1.8015	Other metal ores Mining	6.8904	Commonly used ferrous metals smelting
3	Commonly used ferrous metals smelting	Cement, lime, gypsum manufacturing	1.7511	Tungsten and molybdenum mining and dressing	5.8217	Chinese patent medicines Manufacture
4	Other metal ores Mining	Precious metal rolling Process	1.7341	Commonly used ferrous metals smelting	3.7567	Copper mining and Dressing
5	Copper mining and Dressing	Iron ore mining	1.5105	Cement, lime, gypsum manufacturing	3.5896	Other metal ores Mining
6	Cement, lime, gypsum manufacturing	Graphite and other non-metallic mineral products	1.4830	Chinese patent medicines Manufacture	3.2476	Cement, lime, gypsum manufacturing
7	Inorganic salts Manufacture	Commonly used ferrous metals smelting	0.6266	Copper mining and Dressing	1.7245	Inorganic salts Manufacture
8	Gold mining and dressing	Inorganic acid manufacturing	0.3013	Gold mining and dressing	0.9435	Gold mining and dressing
9	Coal mining and washing	Gold mining and dressing	0.0702	Inorganic salts manufacturing	0.4956	Rare and rare earth metal smelting
10	Lead and zinc Mining	Livestock and poultry slaughter	-0.0598	Graphite and other non-metallic mineral products	0.3556	Lead and zinc Mining

output value, total profits, total profits and taxes, mainly reflecting the economic benefits of industrial industry, so named for the effectiveness factor ( $f_2$ ).

**Factor Analysis and comparison of the improved model ranking:** We calculate the 46 listed-scale industries in Shangluo City of scores on these two factors by using the regression method and average score of the two factors by the weight of each factor contribution rate, so we establish a comprehensive evaluation model as follows:

$$F = 0.577f_1 + 0.423f_2$$

where,  $F = \sum_{i=1}^2 \omega_i f_i$ ,  $f_i$  is common factor,  $w_i = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k}$  is the contribution rate of the common factor (where  $p = 2$ ). We calculate the composite score of 10 industrial, sort of concrete results in Table 3.

Table 3 shows that the current scale factor ( $f_1$ ) scores of the top ten industries are: iron ore mining, manufacturing of proprietary Chinese medicines, which commonly use non-ferrous metal smelting and so on. It shows that the industry is a leading industry of Shangluo City. According to the efficiency factor ( $f_2$ ), the top ten industry are: tungsten and molybdenum mining and other metal mining, cement, lime, gypsum manufacturing, precious metal rolling processing, iron mining and dressing etc. It can be seen that Shangluo City's industrial efficiency mainly depends on a number of energy industry and other industrial economic efficiency is low.

From the comprehensive factor ( $F$ ) of the top 10 industrial sectors, it can be seen that the industrial products market in Shangluo City over a single, mainly depends on its energy and focuses on the energy extraction and processing industries. Like light industry, textile, electronics, handicrafts and other higher value-added products don't have the market advantage.

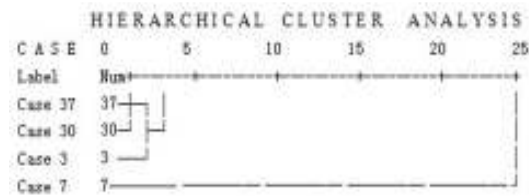


Fig. 1: Shangluo industrial industry cluster analysis pedigree chart

From the improved model (1), it can be seen that the top ten major industries have the same ranking except for the ninth ranking of the scale factor ( $f_1$ ). But the factor analysis shows that the scale factor ( $f_1$ ) in the industry and ranked some of the same, but the gap is far more to improve the model (1)(E) larger. It can be seen from Table 1 that the first two principal components are the cumulative contribution rate of 91.359%, while the first principal component accounted for 83.581%. Various sectors in the original data on the corresponding index value gap is relatively large, the improved model (1) added to reflect this gap between the coefficient of variation, the evaluation results more in line with the actual situation and the scale factor in the ranking the degree of agreement on the improved model (1), the evaluation results are more reasonable.

**Shangluo city industrial structure cluster analysis and evaluation:** Based on the above analysis, we get the Shangluo Industrial industry cluster analysis pedigree chart Fig. 1 by using SPSS two principal components of the selected system clustering.

Shangluo City 46 listed scale industries can be divided into three categories, the third class includes tungsten and molybdenum mine mining; the second class has the iron ore mining, copper mining, metal mining, manufacturing of proprietary Chinese medicines, cement, lime, gypsum manufacturing, which commonly use non-ferrous metal smelting; the others

belong to the first class classification. From the classification, the third class, second class industries of the top seven industries concentrated in the rankings of the comprehensive factor (F) and model (1)(E). These industries ranked in the top ten rankings of the scale factor and the efficiency factor. As a result, we can also obtain that Shangluo City of industrial development is also largely dependent on mineral resources and mostly in the lower level. It is likewise reflected that the scientific nature of the clustering analysis, thus to avoid the artificial classification based on the size of the consolidated scores of subjective and arbitrary. Therefore, the rationality of the model is reflected (1).

### CONCLUSION

In this study, the problems of the principal component analysis were analyzed and discussed from the information provided by the data itself. The results show that we should first conduct a preliminary analysis of raw data in the evaluation of specific issues; we cannot simply extract a number of principal components when the first principal component contribution rate isn't achieved postulate and each sample difference is large. We can choose to rotate the factor loading matrix, select the number of principal components and a combination of factor analysis, coefficient of variation and the variance contribution rate of synthesis, weighted as the weight coefficient, the establishment of a comprehensive evaluation model.

As an example of the core indicators of Shangluo City Industrial Data, a comparative study is conducted based on the composite score by using the factor analysis and the improvement of the model (1), respectively. It is determined that the improvement of the rationality of the model by the classification of Shangluo City Industrial by using cluster analysis methods. The evaluation of analytical results is consistent with the industrial city of Shangluo industrial structure of the existing situation. It provides a reference basis for the further development of Shangluo City Industrial.

### REFERENCES

Gao, C. and N. Cai, 2004. A research on usage of principal component analysis in synthetic ordering with highly correlative variable. *J. Math. Pract. Theory*, 12: 21-24.

- Hu, Y. and S. He, 2000. *Comprehensive Evaluation Method*. Science Press, Beijing, pp: 53-78.
- Li, F. and X. Liu, 2010. The improvement and application about general evaluation of principal component analysis model. *J. Stat. Decis.*, 2: 162-163.
- Lu, W., 2007. *SPSS for Windows Statistical Analysis*. Publishing House of Electronics Industry, Beijing, pp: 477-493.
- Shangluo Statistical Yearbook, 2009. Shangluo Municipal Statistics Bureau. Shangluo, pp: 240-273.
- Shi, W., 2007. Principal component analysis and factor analysis in the practice application of. *J. Stat. Consult.*, 4: 29-33.
- Sun, L. and W. Qian, 2009. An improved method based on principal component analysis for the comprehensive evaluation. *J. Math. Pract. Theory*, 18: 15-20.
- Wang, W. and X. Chen, 2006. Comparison of principal component analysis with factor analysis in comprehensive multi-indicators scoring. *J. Stat. Inform. Forum*, 9: 19-22.
- Xu, Y. and Y. Wang, 2006. The improvement of the application method of principle component analysis. *J. Math. Pract. Theory*, 6: 68-75.
- Yan, C., 1998. On composite evaluation by principal component analysis. *J. Appl. Stat. Manag.*, 3: 22-25.
- Ye, S., 2001. The application and consideration about principal component analysis. *J. Appl. Stat. Manag.*, 2: 52-61.
- Ye, Z., 2006. On issues of comprehensive evaluation by principal component analysis. *J. Zhejiang Normal Univ.*, 6: 87-90.