Asian Agricultural Research 2014, 6(4): 24 – 28

Statistical Analysis and Evaluation of the Economic Development in Shaanxi Province

Xu GUO*, **Chunxiang ZHAO**, **Xinrong YAN**

College of Science, Engineering University of CAPF, Xi'an 710086, China

Abstract Based on 8 indicators of economic development and according to the economic situation of cities in Shaanxi Province, this paper analyzes the data using PCA (Principle Component Analysis) and FA (Factor Analysis) with the help of R software. Using the test of Kendal's W coefficient of concordance, the consistency of results from both methods is tested. Finally, Cluster Analysis is used to classify the results and some advices are proposed for the development of cites in Shaanxi Province.

Key words Principle component analysis, Factor analysis, Kendall's W coefficient

Shaanxi Province is located in the eastern part of Northwest China. As the "bridgehead" of Development of the West Regions, it assumes the role in connecting east and west, south and north, and it has a very important strategic position. However, due to the differences in geographical environment and the distribution of resources, there is disparity in the regional economic development, so the statistical analysis and comprehensive evaluation of the economic development in 10 regions of Shaanxi Province, it is of great significance to formulating the macroeconomic policies and even the development of the entire Northwest region. By reading the literature[1-3], it is not difficult to find that the evaluation results of the same region are not quite similar using the different evaluation methods. In order to overcome the one-sidedness of single evaluation method, this article uses principal component analysis and factor analysis to carry out comprehensive evaluation of regional economy in Shaanxi Province, so the evaluation results obtained will be more objective and reliable.

1 Selection of evaluation indicators

In accordance with the principles of scientificity, systematicness, typicality and operability, we select the following 8 indicators influencing the economic development of various cities in Shaanxi Province from the economic development scale, the economic development structure and the economic development benefits[4]: the total GDP (10^8 yuan), the total retail sales of social consumer goods (10^8 yuan), total imports and exports (10^4 \$), the output value of secondary industry (10^8 yuan), the output value of tertiary industry (10^8 yuan), investment in fixed assets (10^8 yuan), the local revenue (10^8 yuan), and per capita GDP (yuan).

The data for empirical analysis in this paper are from Regional Statistical Yearbook of Shaanxi Province in 2012[5], and some data are calculated. The relevant data are shown in the table below.

Received: December 20, 2013 Accepted: February 25, 2014

Supported by Natural Science Foundation for Fundamental Research in Shaanxi Province (2011JM1019).

* Corresponding author. E-mail: 395849660@qq.com

2 Empirical analysis

2.1 Principal component analysis Principal component analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables.

The steps of principal component analysis are as follows:

(i) Standardizing the original data;

(ii) Calculating the sample correlation coefficient matrix;

(iii) Calculating the characteristic root, variance contribution rate and the eigenvector;

(iv) Selecting the principal component;

(v) Calculating the comprehensive evaluation value based on multi-indicator and weighted comprehensive evaluation model.

By standardizing the original data, we get the standardized matrix B (omitted), calculate the correlation coefficient matrix (omitted), and calculate the eigenvalues of the correlation matrix, variance contribution rate and eigenvector[6] (Table 2).

From the above table, it is found that when the number of principal component is 2, the cumulative contribution rate has reached 98.2% (>85%), including the vast majority of the information needed, and two principal components have been enough. So here we take the first two as the first principal component and second principal component, respectively.

Table 1 Raw data

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Xi'an	3862.6	1 965.98	1257852	1674.31	2015.13	3346.26	318.55	45 475
Tongchuan	232.63	54.56	1017	147.41	67.81	145.96	18.89	27 806
Baoji	1 175.75	358.23	82 268	749.25	297.94	1 008.03	51.44	31 579
Xianyang	1 361.32	345.1	46 641	740.4	368.46	1 263.1	58.56	27 751
Weinan	1 028.97	278.96	17 390	545.19	323.31	912.94	44.3	19 424
Shangluo	362.95	92.18	33 666	163.03	129.31	308.36	16.49	15 513
Ankang	407.17	129.72	1 740	183.13	152.03	304.49	17.34	15 477
Hanzhong	647.48	184.4	4 595	267.58	237.61	411.33	23.55	18 952
Yulin	2 292.25	240.95	10 048	1 629.66	550.68	1378.73	180.25	68 358
Yan'an	1 113.35	131.14	3684	815.45	211.21	815.21	120.73	50 807

Table 2 The loading matrix and variance contribution rate of principal component analysis

Indicators	Principle components F_1	Principle components F_2	Eigenvalues	Contribution rate // %	Cumulative contribution rate // %
X_1	0.383	0.068	6.688	0.836	0.836
X_2	0.357	−0.349	1.171	0.146	0.982
X_3	0.347	−0.384			
X_4	0.347	0.380			
X_5	0.374	−0.232			
X_6	0.380	−0.088			
X_7	0.379	0.130			
X_8	0.240	0.710			

Let Y_1 and Y_2 represent the first principal component and second principal component, respectively, then:

$$Y_1 = 0.383X_1 + 0.357X_2 + 0.347X_3 + 0.347X_4 + 0.374X_5 + 0.380X_6 + 0.379X_7 + 0.240X_8$$

$$Y_2 = 0.068X - 0.349X_2 - 0.384X_3 + 0.380X_4 - 0.232X_5 - 0.088X_6 + 0.130X_7 + 0.710X_8$$

According to the variance contribution rate of the principal component (α_1, α_2) (Table 2), the expression of composite score of the city is as follows: $Y = \beta_1 Y_1 + \beta_2 Y_2$, where $\beta_i = \frac{\alpha_i}{\alpha_1 + \alpha_2}$. The results are shown in Table 3.

Table 3 Principal component score and composite score

	Principal component 1	Principal component 2	Composite score	Ranking
Xi'an	6.579 9	−1.259 5	5.411 8	1
Tongchuan	−1.918 5	−0.119 5	−1.650 4	8
Baoji	−0.278 9	0.096 4	−0.222 9	5
Xianyang	−0.134 0	−0.053 3	−0.122 0	4
Weinan	−0.778 4	−0.440 0	−0.727 9	6
Shangluo	−1.882 1	−0.697 5	−1.705 6	10
Ankang	−1.843 0	−0.682 3	−1.670 0	9
Hanzhong	−1.499 5	−0.543 2	−1.357 0	7
Yulin	1.830 2	2.420 9	1.918 2	2
Yan'an	−0.075 9	1.2779	0.1258	3

2.2 Factor analysis Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. For example, it is possible that variations in four observed variables mainly reflect the variations in two unobserved variables. Factor analysis searches for such joint variations in response to unobserved latent variables.

The observed variables are modelled as linear combinations of the potential factors, plus "error" terms. The information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset. Computationally this technique is equivalent to low rank approximation of the matrix of observed variables. Factor analysis originated in psychometrics, and is used in behavioral sciences, social sciences, marketing, product management, operations research, and other applied sciences that deal with large quantities of data.

Its analysis steps are:

(i) Standardizing the raw data;

(ii) Calculating the correlation coefficient matrix of variables;

(iii) Calculating the eigenvalues, variance contribution rate and eigenvectors;

(iv) Estimating the factor loading matrix;

(v) Building factor score model to derive sample score.

When conducting factor analysis, we need to first carry out KMO test and Bartlett's test of sphericity on data, as shown in Table 4.

Table 4 KMO and Bartlett's test of sphericity

KMO test value		0.711
Bartlett's test of sphericity	Test value	169.616
	Degree of freedom	28
	P value	0

Usually, when KMO test value < 0.5 or Bartlett test Sig value > 0.010, it is not suitable for factor analysis. From Table 4, KMO statistic is 0.711, and Bartlett test Sig value is 0, indicating that the data are suitable for factor analysis.

After the standardization of raw data, we get standardized matrix B (omitted), correlation coefficient matrix (omitted), and calculate the eigenvalues, variance contribution rate of correlation coefficient matrix and factor loading matrix, as shown in Table 5.

From Table 5, it is found that the cumulative variance contribution rate of the first two eigenvalues reaches 96.9%, providing most of the information of the original indicators, so it is appropriate to select the first two factors.

Table 5 Factor loading matrix and variance contribution rate

Indicators	Principle components		Eigen-values	Contribution rate//%	Cumulative contribution rate//%
	F_1	F_2			
X_1	0.759	0.648	4.856	0.607	0.607
X_2	0.972	0.225	2.895	0.362	0.969
X_3	0.982	0.153			
X_4	0.490	0.869			
X_5	0.934	0.351			
X_6	0.845	0.514			
X_7	0.732	0.642			
X_8	0	0.920			

Since the factor loading matrix on the original variables is not easy to explain, we use the orthogonal rotation with the maximum variance to rotate the factor loading matrix.

The expression of factor scores can be obtained as follows according to Table 5:

$F_1 = 0.759X_1 + 0.972X_2 + 0.982X_3 + 0.490X_4 + 0.934X_5 + 0.845X_6 + 0.732X_7$

$F_2 = 0.648X_1 + 0.225X_2 + 0.153X_3 + 0.869X_4 + 0.351X_5 + 0.514X_6 + 0.642X_7 + 0.92X_8$

Then according to the variance contribution rate of the principal component (α_1, α_2) (Table 5), the expression of composite score of the city is as follows: $F = \beta_1 F_1 + \beta_2 F_2$, where $\beta_i = \frac{\alpha_i}{\alpha_1 + \alpha_2}$. The results are shown in Table 6.

Table 6 Factor scores and composite scores

	Principal component 1	Principal component 2	Composite score	Ranking
Xi'an	14.6309	8.615 8	12.381 3	1
Tongchuan	−3.949 7	−2.968 2	−3.582 6	10
Baoji	−0.653 2	−0.310 1	−0.524 9	5
Xianyang	−0.244 5	−0.203 4	−0.229 2	3
Weinan	−1.327 5	−1.487 4	−1.387 3	6
Shangluo	−3.470 1	−3.373 3	−3.433 9	9
Ankang	−3.398 1	−3.298 6	−3.360 9	8
Hanzhong	−2.775 2	−2.670 3	−2.736 0	7
Yulin	2.209 8	4.757 0	3.162 5	2
Yan'an	−1.022 4	0.938 6	−0.289 0	4

2.3 Testing of the results The cities of Shaanxi Province are ranked in terms of the economic development level using two methods. In order to examine whether the analysis results of the two methods are consistent, now we use multivariate Kendall's W Concord coefficient test method to test[7].

The following hypotheses are proposed:

H_0: The analysis results of the two methods are inconsistent;

H_1: The analysis results of the two methods are consistent.

Statistic is as follows:

$$W = \frac{12S}{K^2(N^3 - N)}$$

where S is the quadratic sum of the difference of rank-sum (R_j) of evaluated sample j under k evaluation methods, and the average values.

Namely:

$$S = \sum_{j=1}^{N}[R_j - (\sum_{j=1}^{N} R_j)/N]^2;\ R_j = \sum_{i=1}^{k} R_{ij}$$

where R_{ij} is the rank of sample j in evaluation method i; k is the number of samples to be evaluated.

Here $k = 2$, $j = 10$. R software is used for programming: $W = 0.978$, and significance degree is $0.002 < 0.05$, therefore, it rejects the null hypothesis, and the analysis results of the two methods are believed to be consistent.

The arithmetic mean of scores of the two methods is calculated, and the final score and ranking of the cities are obtained, as shown in Table 7.

Table 7 Composite score and ranking

	Composite score	Ranking
Xi'an	8.896 55	1
Tongchuan	−2.616 5	10
Baoji	−0.373 9	5
Xianyang	−0.175 6	4
Weinan	−1.057 6	6
Shangluo	−2.569 75	9
Ankang	−2.515 45	8
Hanzhong	−2.046 5	7
Yulin	2.540 35	2
Yan'an	−0.081 6	3

2.4 Cluster analysis Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics. Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions.

Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data preprocessing and model parameters until the result achieves the desired properties.

According to the composite scores of the cities in Table 7, the ten cities are clustered, and the results are shown in Fig. 1.

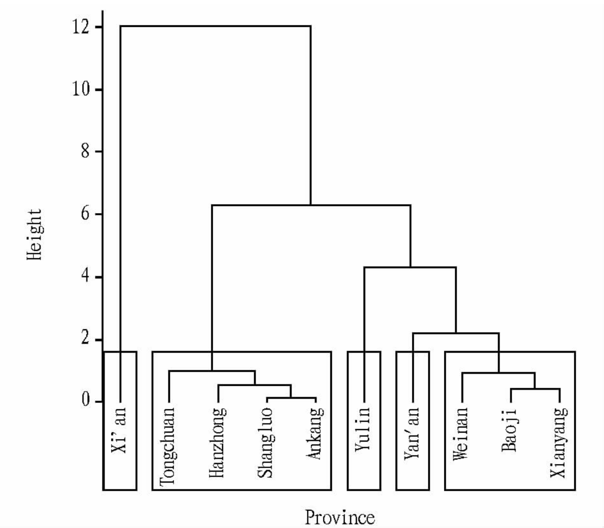

Fig. 1 Clustering results

According to the composite scores of the cities, Shaanxi Province can be divided into five categories:

(i) Xi'an falls within one category;

(ii) Yulin falls within one category;

(iii) Yan'an falls within category;

(iv) Weinan, Baoji and Xianyang fall within one category;

(v) Tongchuan, Hanzhong, Shangluo and Ankang fall within one category.

As can be seen from the system clustering diagram, there are great differences in the level of economic development between the cities of Shaanxi Province. Xi'an is developed region; Yulin, Yan'an, Weinan, Xianyang, and Baoji are comparatively developed regions; Tongchuan, Hanzhong, Ankang and Shangluo are underdeveloped regions.

Classification category has obvious geographical similarity, that is, there is a close relationship between the degree of development of these cities and their location.

3 Conclusions and recommendations

According to the composite scores and clustering analysis results, coupled with the study results of previous scholars[9], we can see that there are strong geographical features in the economic development of cities in Shaanxi Province: Xi'an has the highest level of economic development, followed by the northern regions of Shanxi Province; other cities in the Guanzhong area are in the middle ranks; the southern regions of Shanxi Province have the lowest level of economic development. So Shaanxi Province can be roughly divided into four gradient regions of economic development.

(i) The first gradient developed region: Xi'an. Xi'an is the capital of Shaanxi province, and a sub-provincial city in the People's Republic of China. One of the oldest cities in China, with more than 3100 years of history, the city was known as Chang'an before the Ming Dynasty. Xi'an is one of the Four Great Ancient Capitals of China, having held the position under several of the most important dynasties in Chinese history, including Zhou, Qin, Han, Sui, and Tang. Xi'an is the starting point of the Silk Road and home to the Terracotta Army of Emperor Qin Shi Huang.

It is a tourist city, and the development of its tertiary industry ranks first in the province; its economic strength is also second to none in Shaanxi Province and even the whole western region. In terms of the composite score, the gap between it and the second place city Yulin is also very significant, therefore, Xi'an should strengthen communication and contact with the surrounding cities and drive their economic development while maintaining its rapid development.

(ii) The second gradient comparatively regions: Yulin and Yan'an. Northern Shaanxi's economic development is rapid in recent years, especially for Yulin City. In the context of Development of the West Regions, these regions have given full play to their own advantages, and used rich mineral resources and developed industry to drive the economic development of the entire region.

At the same time, they take full advantage of the unique ethnic customs and tourism resources, to promote the rapid development of the tertiary industry in northern Shaanxi region. By virtue of good economic development model, Yan'an City has been consistently ranked in the forefront of Shaanxi Province.

(iii) The third gradient regions with stable development: Baoji, Xianyang and Weinan. These three cities are located in the Guanzhong area, with better opportunities for development. "Xi'an – Xianyang Integration" is a new opportunity for the development of Xianyang City. Weinan City is a city with moderate level of economic development in the Guanzhong area.

Its tourism resources are abundant, but its level of industrial development is relatively backward, so the overall economic level is still relatively low.

(iv) The fourth gradient underdeveloped regions: Tongchuan, Hanzhong, Ankang and Shangluo. In these regions, the economic foundation is weak, and the industrial development is relatively backward, but there is a wealth of water resources, mineral resources and biological resources.

It is necessary to strengthen exchanges and cooperation with other economic zones, use the superior resources to develop industry, and take a suitable economic model.

References

[1] [1] ZHANG XT. Comprehensive evaluation on economic development level in Shaanxi Province[J]. Shaanxi Electric Power, 2007, 35(5): 77 – 80. (in Chinese).

[2] YANG XP, WU PZ. The regional development level spatial differential studying of Shaanxi[J]. Journal of Gansu Lianhe University(Natural Sciences), 2007, 21(5): 59 – 62. (in Chinese).

[3] DIAO Y, WANG YH. Analysis on economic development of cities in Shaanxi Province based on factor analysis[J]. Industrial Technology & Economy, 2009, 28(9): 116 – 119. (in Chinese).

[4] LIU JS. Synthetical strength analyses and evaluation for cities in Jiangxi [J]. Prices Monthly, 2010(1): 27 – 29. (in Chinese).

[5] Shaanxi Bureau of Statistics (NBS). West Shaanxi statistical yearbook, 2012 [M]. Shaanxi Statistical Publishing House, 2012. (in Chinese).

[6] XUE Y, CHEN LP. R software modeling and R software[M]. Beijing: Tsinghua University Press, 2007. (in Chinese).
[7] WEN YY, HU P. A comprehensive application of several evaluation methods to the investment environment evaluation [J]. Economic Geography, 2002, 22(4): 390 - 393. (in Chinese).
[8] XING HH, LIU KW. A changing analysis of the comprehensive economic strengths of Shaanxi Province[J]. Ecological Economy, 2007(6): 65 - 68. (in Chinese).

(From page 23)

japonica. Bureau of science and technology in Chuzhou city placed the introduction, experiment and demonstration of new varieties of japonica and new technology as well as the research of matching technologies on the list of key research projects. Agricultural committee took the work of transforming indica to japonica as the priority, strengthened service and supervision as well as widely carried out the construction of demonstration areas. Agricultural machinery bureau enhanced the technical integration of mechanized transplanting and matching service as well as quickened the promotion of mechanization of japonica rice production. Other agriculture-related departments like water conservancy and land resources enhanced the construction of agricultural infrastructure, strengthened propagation service and promptly sent key information like agricultural technology, weather as well as supply and demand to major households through agricultural information platform. Furthermore, cooperatives and specialized serving teams have developed quickly and provided matching service, such as seedling cultivation, mechanized farming, mechanized transplanting, mechanized preservation and reaping for those scaled rice farmers, which has promoted the dramatic development of japonica rice production.

References

[1] SHI SS, QI YL, LI CL, et al. Primary discussions on changing Jopnica to Indica for planting in the south of Henan Province[J]. Tillage and Cultivation, 2007(2): 40 - 41. (in Chinese).