

Comprehensive Evaluation and Study on Regional Economic Development of Huangshi City

¹Yazhou Xiong and ²Qin Li

¹Department of Economics and Management,

²Normal Department, Hubei Polytechnic University, Huangshi, 435003, China

Abstract: Under the background of “Two-oriented Society”, it was one of the important content of promoting the construction and development of Wuhan urban agglomeration to achieve the rapid development of the deputy center city (Huangshi). In this study, statistical data, the principal component analysis and K-means clustering algorithm are integrated to assess and analyze the economic development of Huangshi City in Wuhan urban agglomeration in 2010. Finally this analysis is applied to real data, it is demonstrated that the result are clear and realizable, which shows that proposed evaluation method can provide enough instructions for the real cases.

Keywords: Huangshi, k-means clustering algorithm, principal compound analysis, regional economy

INTRODUCTION

Wuhan urban agglomeration is also known as the "1+8" urban agglomeration, in which the capital of Hubei province: Wuhan city is the core city and surrounded by other eight cities including Huangshi, Huanggang, Xiaogan, Xianning, Ezhou, Xiantao, Tianmen and Qianjiang. As deputy core city of Wuhan urban agglomeration, Huangshi is an important base of raw material industry in Central China since more than 100 years ago and it is also one of the cities to open to the outside world along the Changjiang River which have been approved by the Chinese State Council. Its coordinated growth of the regional economy of Huangshi has very important significance under the background of considering "two-oriented society". At present, there have been many researches on regional development and many scholars have adopted lots of methods, such as qualitative analysis, Principal Component Analysis (PCA), econometric models (Xia, 2007; Li *et al.*, 2009). But when Huangshi was discussed, most of the current studies emphasized on it's the qualitative aspect of the region economy (Meng, 2010), which means there is no quantitative studies about Huangshi now. In this study, the principal component analysis is employed to select a few typical, representative indicators from many indexes related to regional economic development and the economic data of Huangshi in 2010 is analyzed. From the results it is found that the observation index is improved and simplified, which means it can comprehensively analyze and evaluate the regional economic growth of Huangshi city.

The general situation of the study area: Huangshi city is in the southeast of Hubei Province in China and located the south bank of the Changjiang River. It has a long history of mining, rich in cultural heritage, a solid industrial foundation, convenient geographical location and very beautiful natural environment. The total area of this city is 4583 km² and the population of resident at the end of the year of 2010 reached 243 million; GDP arrived to 69.012 billion Yuan (distribution of three industries is 7.77, 57.22 and 35.01%, respectively). The administrative division is divided into six parts, including Daye City, Yangxin County, Huangshigang District, Xisaishan District, Xialu District and Tieshan District. Due to geographical location, climate conditions, resources, economic foundation and other factors, there are unbalanced development of regional economy of Huangshi city. Therefore, it is extremely significant to analyze and evaluate these differences, which is helpful for realizing the coordinated development of different regions.

THE EVALUATION INDEX SYSTEM

Considering the available data and scientific and comprehensive evaluation index, the author establishes the evaluation index system of regional economic development region development of Huangshi city from three aspects, which include the total economy, the economic structure and the people's living standards (Li *et al.*, 2011). This system is consisted of three first-grade indicators, thirteen second-grade indicators are given, which are shown as Table 1.

Table1: The evaluation index system

First-grade indicators	Second-grade indicators	Unit	No
Total economy	Regional GDP	Billion Yuan	W ₁
	Total caliber local revenue caliber	Billion Yuan	W ₂
	Total investment in fixed assets	Billion Yuan	W ₃
	Industrial added value above designated size	Billion Yuan	W ₄
	Total social retail sales	Billion Yuan	W ₅
	Gross foreign export value	Million dollars	W ₆
Economic structure	The primary industry proportion	%	W ₇
	The secondary industry proportion	%	W ₈
	Tertiary industry proportion	%	W ₉
People's living standards	Total population (resident)	10 ⁴	W ₁₀
	The natural population growth rate	%	W ₁₁
	The average per capita disposable income of urban residents	Yuan	W ₁₂
	Number of new jobs in cities and towns	People	W ₁₃

COMPREHENSIVE EVALUATION MODEL

Principal component analysis is a multivariate statistical method which can convert the multiple related variables (index) of the research object into a few irrelevant variables. The evaluation index system of regional economy development is a multi-level integrated system, in which many other variables are involved and also there are interactive relationships among these indexes. Therefore, there is a certain degree of overlapping in the information that the statistical data reflects. In order to eliminate overlapped information, in this study the principal component analysis is adopted. On the basis of retaining most of original variables information, several principal components which are not related to each other are used as linear combinations of the many original variables. In this way it can reduce corresponding computational effort and provide more concise comprehensive evaluation (Zhang *et al.*, 2004).

To construct the sample matrix X: Suppose the number of evaluation object is n, the number of evaluation index is p and thus given sample values constitute the sample matrix:

$$X_i = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{ij}\} \tag{1}$$

where, i = 1, 2, 3, ..., n; j = 1, 2, 3, ..., p

To convert the element of the sample matrix X:

$$t_{ij} = \begin{cases} x_{ij}, & \text{On the positive index} \\ -x_{ij}, & \text{On the negative index} \end{cases}$$

Then:

$$T = [t_{ij}]_{n \times p} \tag{2}$$

Correlation coefficient matrix R: After the standardized transformation of matrix elements,

standard matrix Z can be obtained and then the correlation coefficient matrix can also be obtained through Z:

$$R = [r_{ij}]_{p \times p} = \frac{Z^T Z}{n - 1} \tag{3}$$

where,

$$r_{ij} = \frac{\sum_{k=1}^n z_{kj} z_{ki}}{n - 1}, i, j = 1, 2, \dots, p$$

To find eigen values: After solving the characteristic equation of the sample correlation coefficient matrix, we can get the corresponding the Eigen value:

$$|R - \lambda I_p| = 0 \tag{4}$$

Then

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$$

where,

R = Correlation coefficient matrix

λ_i = The Eigen value; (i = 1, 2, 3, ..., p);

To determine the main components:

$$\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.85 \tag{5}$$

According to the Eq. (5), m principal components can be determined, which will absorb 85% of the utilized information.

To solve the unit feature vector W:

$$\{y_1, y_2, \dots, y_m\} = Z \times W \tag{6}$$

Table 2: The sample data table in 2010

Regions	W ₁	W ₂	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉	W ₁₀	W ₁₁	W ₁₂	W ₁₃
Huangshigang	84.800	5.69	38.30	5.530	81.50	3009	0.550	29.75	69.69	22.43	2.63	16658	9290
Xisaishan	102.37	7.83	55.74	11.07	51.00	25000	1.130	68.10	30.77	23.37	2.15	14153	8100
Xialu	78.550	5.35	38.66	18.24	16.00	4003	0.500	79.31	20.19	17.66	5.10	15083	4452
Tieshan	26.200	2.52	18.70	5.580	9.120	6618	0.690	78.71	20.60	5.730	1.88	14967	4545
Daye	251.08	25.95	152.59	118.49	80.57	3538	9.790	60.00	30.21	90.97	4.50	15020	10800
Yangxin	110.71	7.21	93.56	26.51	58.45	1502	23.76	36.41	39.83	82.20	10.19	11028	7859

Data derived from the "2011 statistical yearbook of Huangshi" and the website of provincial bureau of statistics

Table 3: Correlation matrix

	ZW ₁	ZW ₂	ZW ₃	ZW ₄	ZW ₅	ZW ₆	ZW ₇	ZW ₈	ZW ₉	ZW ₁₀	ZW ₁₁	ZW ₁₂	ZW ₁₃
ZW ₁	1.00	0.98	0.96	0.95	0.68	-0.12	0.38	-0.19	0.02	0.83	0.23	-0.05	0.79
ZW ₂	0.98	1.00	0.93	0.98	0.60	-0.12	0.28	-0.07	-0.07	0.76	0.10	0.04	0.73
ZW ₃	0.96	0.93	1.00	0.92	0.64	-0.20	0.61	-0.25	-0.03	0.94	0.44	-0.31	0.74
ZW ₄	0.95	0.98	0.92	1.00	0.49	-0.25	0.34	0.00	-0.17	0.78	0.19	-0.02	0.62
ZW ₅	0.68	0.60	0.64	0.49	1.00	-0.10	0.34	-0.78	0.73	0.62	0.15	0.05	0.97
ZW ₆	-0.12	-0.12	-0.20	-0.25	-0.10	1.00	-0.38	0.36	-0.22	-0.35	-0.49	0.01	0.01
ZW ₇	0.38	0.28	0.61	0.34	0.34	-0.38	1.00	-0.50	0.07	0.83	0.91	-0.85	0.32
ZW ₈	-0.19	-0.07	-0.25	0.00	-0.78	0.36	-0.50	1.00	-0.90	-0.43	-0.41	0.17	-0.63
ZW ₉	0.02	-0.07	-0.03	-0.17	0.73	-0.22	0.07	-0.90	1.00	0.07	0.01	0.24	0.57
ZW ₁₀	0.83	0.76	0.94	0.78	0.62	-0.35	0.83	-0.43	0.07	1.00	0.69	-0.53	0.67
ZW ₁₁	0.23	0.10	0.44	0.19	0.15	-0.49	0.91	-0.41	0.01	0.69	1.00	-0.82	0.08
ZW ₁₂	-0.05	0.04	-0.31	-0.02	0.05	0.01	-0.85	0.17	0.24	-0.53	-0.82	1.00	0.04
ZW ₁₃	0.79	0.73	0.74	0.62	0.97	0.01	0.32	-0.63	0.57	0.67	0.08	0.04	1.00

Table 4: Total variance explained

Component	Initial eigen values			Extraction S.S. loadings		
	Total	(%) of variance	Cumulative (%)	Total	(%) of variance	Cumulative (%)
1	6.619	50.917	50.917	6.619	50.917	50.917
2	2.728	20.983	71.900	2.728	20.983	71.900
3	2.524	19.413	91.314	2.524	19.413	91.314
4	1.040	8.0020	99.316			
5	0.089	0.6840	100.000			

The total of components is 13, and the other 8 principal component values are very small, so they do not been listed

where,

y_i = The principal component linear combination of λ_i

$i = 1, 2, 3, \dots, m$

Z = Standardized matrix

W = Unit feature vector

After obtaining the former m Eigen values and corresponding feature vector, we must make them standardized and receive the unit feature vector W. According to the Eq. (6), we can get the principal component linear combination.

To perform final evaluation:

$$F = a_1 \hat{y}_1 + a_2 \hat{y}_2 + \dots + a_m \hat{y}_m \quad (7)$$

where,

\hat{y}_i = The score of the principal component i

a_i = The variance weight of principal component i

$i = 1, 2, \dots, m$

F = The score of the comprehensive evaluation

Case studies: In this study, the strategy of integrating principal component analysis and K-means Cluster is employed. And sample date of the six regions of Huangshi city in 2010 are selected as the analysis

object. Based on the principles and procedures of our method, SPSS software is used to process data and makes comprehensive evaluation (Zhu *et al.*, 2010). The sample data are shown in Table 2.

Considering the effect of dimensional differences, data must be standardized before they are analyzed. This study uses SPSS software to analyze the data and makes the original data standardized; finally obtains the correlation coefficient matrix which is illustrated as Table 3. The principal component feature value arranged from largest to smallest values, variance (contribution), the cumulative variance (the cumulative contribution rate) are given in Table 4. As can be seen

Table 5: Unit feature vector

Component	1	2	3
W ₁	0.348	0.181	0.197
W ₂	0.322	0.226	0.260
W ₃	0.370	0.024	0.189
W ₄	0.316	0.152	0.306
W ₅	0.307	0.230	-0.288
W ₆	-0.123	0.183	0.131
W ₇	0.272	-0.420	-0.057
W ₈	-0.206	0.042	0.531
W ₉	0.099	0.166	-0.583
W ₁₀	0.375	-0.143	0.073
W ₁₁	0.206	-0.490	-0.058
W ₁₂	-0.126	0.518	-0.051
W ₁₃	0.322	0.267	-0.165

Table 6: Principal component score, total score and classification

Regions	\hat{F}_1	Rank	\hat{F}_2	Rank	\hat{F}_3	Rank	F	Rank	Classification results
Huangshi-gang	-0.1739	3	1.5352	1	-2.7739	6	-0.3049	4	2
Xisaishan	-1.0173	4	0.847	3	0.3978	4	-0.2631	3	2
Xialu	-1.9226	5	-0.565	5	1.0214	2	-0.8992	5	2
Tieshan	-2.9201	6	-0.2908	4	0.7394	3	-1.4043	6	2
Daye	3.9124	1	1.3576	2	1.5297	1	2.5739	1	1
Yangxin	2.1215	2	-2.884	6	-0.9143	5	0.2975	2	3

\hat{F}_i = The score of the principal component $i(i = 1, 2, 3)$.

from Table 4, the top 3 principal components of the feature values are greater than 1 and the cumulative variance contribution rate has been in excess of 85% and reached 91.314%, which can be basically reflect the overall situation. So this study selects 3 indicators as the main components targets.

Then feature vector corresponding to the top three eigen values can be solved and standardized, which are shown in Table 5. According to the above Eq. (7), each principal component linear combination can be obtained. Obviously, the loads of the first principal component in $W_1, W_2, W_3, W_4, W_5, W_{10}, W_{13}$ is larger, which mainly reflects the level of economic development from the economic gross; the loads of the second principal component in W_7, W_{11}, W_{12} is larger, which mainly reflects the level of economic development from the average per capita disposable income of urban residents and the first industry accounted for the proportion of GDP and the feedback of the W_7, W_{11} is negative; the loads of the third principal components in W_8, W_9, W_4 is larger, which mainly reflects the level of economic development from the second industry accounted for the proportion of GDP, above-scale industrial added value and so on and the feedback of the W_9 is negative.

The principal component scores are comprehensively weighted according to their respective contribution rate and final evaluation will be concluded. Through the Table 4, the contribution rate of three main components is obtained as 50.917, 20.983 and 19.413%, respectively. According to the Eq. (6), we input the normalized data and calculate the principal component score of the six regional of Huangshi city in 2010. Finally six regional will be sorted by the general score from the highest to the lowest. As illustrated in Table 6, the higher comprehensive score means higher region economy development level; on the contrary it has lower level. Meanwhile, according to the factor scores of each region, we perform the clustering analysis by the means of K-means clustering method in SPSS software. Before the clustering analysis is conducted, it is supposed that the cluster number is 3 and the convergence criteria value is 0.02, the final classification results can be shown in the last column of Table 6.

DISCUSSION OF RESULTS

According to above statistics data and results of the analysis by means of our system, six regions of Huangshi city are sorted into three levels.

The first level: The first level includes the city of Daye. As it is only county-level city of all regions, it has largest economic aggregate with highest composite score. It has developed industries and the per capita disposable income of urban residents is the second highest in all regions. All these data is consistent with the actual situation.

The second level: The second level includes Huangshigang District, Xisaishan District, Xialu District and Tieshan District. Their situations will be introduced separately:

- **Huangshigang district:** As illustrated in Table 6, the second principal component score of Huangshigang ranks first in all six regions: its per capita disposable income of urban residents is the highest. At the same time, the third principal component score is negative: it ranks last in six regions. The secondary industry proportion is close to 29%, but the service sector accounted for the proportion of nearly 70%, which indicates the tertiary industry of this region has big advantages. This district has been approved for the services comprehensive supporting reform pilot area of Wuhan urban agglomeration as the first batch of modern service industry demonstration zone of Hubei Province. Therefore it should continue to use its advantages and tries to obtain more development of service industry (James *et al.*, 1989).
- **Xisaishan district:** The first principal component score of this district is negative and the second and the third principal component score is not high too. The second industry proportion of GDP is close to 70% and third industry accounted for the proportion of GDP is 30%, which indicates that it should make more efforts to advance the construction of industrial zones, develop variously service sectors and try to speed up industrial restructuring.
- **Xialu district:** The third principal component score is higher; however its first and second principal component scores are negative. Its industrial development is good, but the total economic aggregate is small. The three big traditional mainstay industries including non-ferrous metal, black metal, machinery manufacturing of the area continue to grow up and develop. At the same time, this area has been listed as national new industrialization

demonstration base. Its copper smelting and deep processing industrial cluster is listed as key development industry group in Hubei province. Therefore, for this district it should strengthen its traditional primary industries and accelerate the development of modern service industry.

- **Tieshan district:** The third principal component score is higher and the industry grows well too. But its first principal component score ranks last in six regions. Besides its economic aggregate is small. Thus it should continue to carry out infrastructure construction of manufacturing industrial parks and promote rapid development of industrial park.
- **The third level:** The third level includes Yangxin County. Both of the second and third principal component scores are negative and they are almost the lowest in all the six regions. The average per capita disposable income of urban residents is relatively low and its modern industry is considered as underdeveloped. The majority of industrial enterprises in the county have low capacity of production; product technology content and added value are low; rural infrastructures are weak too. New rural reconstruction here has big unbalance with other districts.

CONCLUSION

In this study, a regional economic development level of indicators system for Huangshi city was established. With this system, principal component analysis and K-means clustering method are combined to analyze the statistics data of this city in 2010. After obtaining three main components it can provide the basic situation and ranking based on the evaluation scheme. Finally compensative evaluation and analysis on the regional development were given, which is instructive and meaningful to local coordinated development of Huangshi city's regional economic.

ACKNOWLEDGMENT

This study was financially supported by Scientific Research Projects of Huangshi Institute of technology (2009yjr77B) and Teaching Research Projects of Huangshi Institute of technology (201130).

REFERENCES

- James, O., R. Wheeler and L. Mitchelson, 1989. Information flows among major metropolitan areas in the United States. *Ann. Assoc. Geograph.*, 79(4): 52-543.
- Li, X. and S. Zhang, 2009. Application of principal components analysis in regional economy analysis. *Comput. Eng. Applications.*, 45(19): 204-206.
- Li, J., Y. Bai, J. Luo and Y. Huang, 2011. Spatial analysis of economic disparities of county level in Gansu province. *Econ. Geography*, 31(3): 390-395.
- Meng, H., 2010. Analysis and countermeasure analysis on county economy development of Huangshi. *J. Hubei Normal Univ. Philosophy Soc. Sci.*, 30(6): 105-107.
- Xia, J., 2007. Policy research of the region economy coordinated development. *Sci. Technol. Prog. Policy.*, 24(12): 40-42.
- Zhang, H. and L. Feng, 2004. Application of the principal component analytical method to S and T innovation capability evaluation of universities. *J. Wuhan Univ. Technol. Inform. Manage.*, 26(6): 157-161.
- Zhu, J., R. Zhong, Y. Yang and Z. Chen, 2010. Research on coordinated development of Wuhan urban agglomeration based on principal component analysis. *Resour. Dev. Market.*, 26(2): 120-124.