Economics & Business Management
ISSN Print: 3079-5214
ISSN Online: 3079-5222
https://doi.org/10.63313/EBM.9134

Research on the Evaluation of Urban Scientific and Technological Innovation Capacity in Anhui Province Based on Factor Analysis and Cluster Analysis

Saihong Huang
Faculty of Statistical and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233000, China
Email: 2672736543@qq.com

How to cite this paper: Huang, S. H. (2025). Research on the Evaluation of Urban Scientific and Technological Innovation Capacity in Anhui Province Based on Factor Analysis and Cluster Analysis. *Economics & Business Management*, 4(1), 120-129. ISSN Print: 3079-5214; ISSN Online: 3079-5222.
https://doi.org/10.63313/EBM.9134
Published: 2025-12-25

Abstract

Based on the three aspects of innovation input, innovation environment and innovation output, this paper selects 12 indicators to construct the evaluation index system of scientific and technological innovation capability of 16 prefecture-level cities in Anhui Province. Firstly, the factor analysis method was used to set the innovation input and output factors and the innovation environment factors, and the ratio of the variance contribution rate of each factor to the cumulative contribution rate was used as the weight to obtain the comprehensive score, and the scientific and technological innovation ability of each city was ranked and analyzed. Then, through the cluster analysis method, the 16 cities in Anhui Province were divided into four categories and analyzed for different types of cities. Finally, it is proposed that cities in Anhui Province should improve innovation policies, strengthen industrial agglomeration, optimize innovation environment, and introduce innovative talents, so as to maximize the innovation potential of cities in Anhui Province and better enhance the overall scientific and technological innovation ability of cities in Anhui Province.

Keywords

Factor analysis; cluster analysis; scientific and technological innovation; Competency evaluation

1. Introduction

Scientific and technological innovation is the primary driving force for high-quality development and a key support for accelerating regional economic transformation. The report of the 19th National Congress of the Communist Party of China clearly stated that innovation is the primary driving force for development. Against this background, the deep integration of the innovation-driven development strategy is particularly critical. In recent years, scholars have conducted in-depth evaluation research on technological innovation capability using diverse methods and indicators, yielding many constructive conclusions and suggestions. These studies mainly focus on two directions. On the one hand, scholars are committed to constructing a comprehensive and accurate evaluation index system for scientific and technological innovation capability. For example, Liu Zhongwen et al. (2009) constructed an evaluation system with 22 elements from multiple dimensions such as technological innovation input, support, diffusion, output, and sustainable innovation [1]. Zhang Xuping et al. (2010) constructed an extensive evaluation system with 40 indicators from four levels: input, output, diffusion, and support [2]. Li Meizhi (2012) and Bei Shuhua et al. (2021) also selected multiple indicators from different perspectives to construct targeted evaluation systems [3]. On the other hand, scholars are constantly exploring research methods for evaluating scientific and technological innovation capability. For example, Zhang Jianwei et al. (2022) conducted a detailed study on the innovation capability and economic development of Henan Province by comprehensively applying coupling coordination models, spatial autocorrelation models, and multiple linear regression models[4]. Sun Lijie et al. (2007) used factor analysis to comprehensively evaluate the technological innovation capability of 12 typical cities in eastern, central, and western China[5]. Wang Jun et al. (2020) combined principal component analysis and cluster analysis to analyze and evaluate the innovation capability of high-tech industries in Anhui Province[6].

Although Yang Suchang et al. (2021) used principal component analysis in the process of constructing an evaluation index system for Henan Province's independent innovation capability, they did not deeply analyze the specific performance behind the scores of each common factor [7]. In view of this, this study aims to construct a more accurate and effective evaluation index system for urban technological innovation in Anhui Province. First, factor analysis is used to extract key factors to simplify the complex evaluation system and ensure the consistency and accuracy of the evaluation. Furthermore, cluster analysis is used to classify and evaluate the 16 cities in Anhui Province in detail.

2. Research Design

2.1. Evaluation System for Urban Scientific and Technological Innovation Capability in Anhui Province

Urban scientific and technological innovation capability is affected by multiple innovation elements and is the result of the comprehensive linkage and feedback of various innovation elements. This paper selects three dimensions—innovation input, innovation environment, and innovation output—to construct the evaluation index system for urban scientific and technological innovation capability in Anhui Province [8], as shown in Table 1.

Table 1. Evaluation Index System of Urban Science and Technology Innovation Ability in Anhui Province

	Evaluative dimension	Evaluating indicator	Unit	Variable
The scientific and technological innovation capacity of Anhui Province	Innovation investment	the proportion of local finance expenditure on science and technology in public finance expenditure	%	X1
		The proportion of R&D expenditure to main business revenue in industrial enterprises	%	X2
		R&D expenditure as a percentage of GDP	%	X3
		R&D personnel equivalent to full-time equivalent	person-years	X4
	Innovative environment	Number of high-tech enterprises	unit	X5
		Number of tech business incubators	unit	X6
		the proportion of enterprises with R&D activities in total enterprises	%	X7
		proportion of enterprises establishing R&D institutions	%	X8
	innovation output	Number of published scientific papers	article	X9
		Number of patents authorized for application	piece	X10
		Sales revenue from new products	10000 Yuan	X11
		number of registered scientific and technological achievements	item	X12

2.2. Sample Data Acquisition and Preprocessing

This research is based on the data disclosed in the "2022 Anhui Statistical Yearbook" and the "Statistical Bulletin on National Economic and Social Development" of each city in 2022. The data of 16 prefecture-level cities under the jurisdiction of Anhui Province in 2022 were selected as the analysis samples, and the data were standardized.

2.3. Research Methods

2.3.1. Factor Analysis Model

Factor analysis is a statistical method used to simplify data structure. It reduces the complexity of data by identifying underlying patterns in the dataset. The method is based on a core assumption: behind many observable variables, there are several latent factors that cannot be directly measured, and these factors play a decisive role in the variation of the variables. Therefore, the main task of factor analysis is to explore these latent factors and clarify their internal relationship with the original variables, thereby transforming complex datasets into a simpler, more understandable, and interpretable form.

The factor model is expressed as Equation (3).

$$\begin{aligned} x_1 &= \mu_1 + \alpha_{11} f_1 + \alpha_{12} f_2 \cdots + \alpha_{1n} f_n + \varepsilon_1 \\ x_2 &= \mu_2 + \alpha_{21} f_2 + \alpha_{22} f_2 \cdots + \alpha_{2n} f_n + \varepsilon_2 \\ &\vdots \\ x_m &= \mu_m + \alpha_{m1} f_2 + \alpha_{m2} f_2 \cdots + \alpha_{mn} f_n + \varepsilon_m \end{aligned} \tag{3}$$

α_{ij} represents the correlation coefficient between the i-th variable and the j-th common factor. It reflects the importance of the i-th variable relative to the j-th common factor. This coefficient is called the factor loading. The symbolic representation of the correlation matrix is shown in Equation (4).

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} F=\begin{bmatrix} f_1 \\ f_2 \\ \cdots \\ f_n \end{bmatrix} \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} \mu=\begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_n \end{bmatrix} \tag{4}$$

The factor analysis model can be expressed in the form of a matrix equation, as shown in Equation (5).

$$X=\mu+AF+\varepsilon \tag{5}$$

In the initial step of factor analysis, the original variable X is standardized so that the new variable has a mean of 0 and a variance of 1. In this case, if the constant term is set to 0, the factor analysis model containing K common factors can be expressed by Equation (6).

$$x_i=\alpha_{i1}f_1+\alpha_{i2}f_2+\cdots+\alpha_{ik}f_k+\varepsilon_i \tag{6}$$

The factor analysis model can be expressed as Equation (7).

$$X=AF+\varepsilon \tag{7}$$

In addition, it should be noted that when performing factor analysis, it is necessary to clarify that n<m. The common factors are uncorrelated with each other, the correlation coefficient of each common factor is zero and the variance is standardized to 1, and each specific factor is also required to be uncorrelated with each other.

2.3.2. Cluster Analysis

Cluster analysis is a statistical technique that groups objects in a dataset. The purpose is to make the data individuals in the same group have great similarity in characteristics, while the characteristics of individuals in different groups are quite different. The average linkage method makes good use of the information between all samples. First, the density of data points is calculated through a density estimation method. Areas with high density usually represent the center of the cluster. Based on the density estimation, through continuous iteration, the data points are "translated" along the direction of the density gradient until they reach the density peak or converge to a stable state. When the data points complete the translation and stabilize, the data points located near the same density peak are classified into the same cluster, forming the clustering result.

3. Empirical Analysis of Urban Scientific and Technological Innovation Capability in Anhui Province

3.1. Evaluation of Urban Scientific and Technological Innovation Capability in Anhui Province Based on Factor Analysis

3.1.1. Applicability Test of Factor Analysis

The applicability of factor analysis can be assessed using the Kaiser-Meyer-Olkin (KMO) test statistic and Bartlett's test. The fundamental principle of factor analysis lies in evaluating the relationship between the intercorrelation of observed variables and their shared variance. When observed variables exhibit high correlation, factor analysis is more likely to yield positive results, suggesting the presence of underlying common factors. As shown in Table 2, the KMO test value for the data used in this study is 0.719, exceeding the 0.6 threshold, indicating suitability for factor analysis. Additionally, the P-value of the Bartlett's test is 0.000, which is below the 0.05 significance level, thereby confirming the applicability and validity of factor analysis.

Table 2. KMO and Bartlett's Test Results

Test		Numeric value
KMO test		0.719
Barlett's sphericity test	Approximate Chi-square	312.71
	Degree of freedom	66
	P value	0.000

3.1.2. Extraction of Common Factors

When extracting common factors, the main criterion to be followed is to minimize the loss of information. This can be achieved through considerations in three aspects: First, factors with large explained variance and eigenvalues greater than 1 should be retained, as these factors are usually of high importance; Second, the retained factors should cover a significant proportion of the total variance, usually with a cumulative variance contribution rate of over 80% as the evaluation standard; Third, a scree plot is drawn and the inflection point of the curve is analyzed. The factors before the inflection point are considered more critical. According to the total variance explanation shown in Table 3, two factors were extracted, with their respective variance contribution rates being 69.16% and 15.89%. The cumulative contribution rate of these two main common factors is slightly higher than 85%, indicating minimal information loss. Furthermore, combined with the scree plot of factor analysis (as shown in Figure 1), it can be confirmed that the extracted common factors are effective. Therefore, it can be considered that the original 12 indicators can be effectively synthesized into two main factors, namely F_1 and F_2.

Table 3. Eigenvalues and variance contribution rates

Ingredent	Initial eigenvalue			Extract load sum of squares			Sum of squares of rotational loads		
	Total	Variance percentage (%)	Cumulative percentage (%)	Total	Variance percentage (%)	Cumulative percentage (%)	Total	Variance percentage (%)	Cumulative percentage (%)
1	8.299	69.160	69.160	8.299	69.160	69.160	8.271	68.920	68.920
2	1.907	15.890	85.050	1.907	15.890	85.050	1.935	16.130	85.050
3	0.666	5.550	90.600						
4	0.627	5.230	95.820						
5	0.241	2.010	97.830						
6	0.128	1.070	98.910						
7	0.074	0.620	99.520						
8	0.037	0.310	99.830						
9	0.012	0.100	99.930						
10	0.006	0.050	99.980						
11	0.002	0.020	100.000						
12	0.001	0.000	100.000						

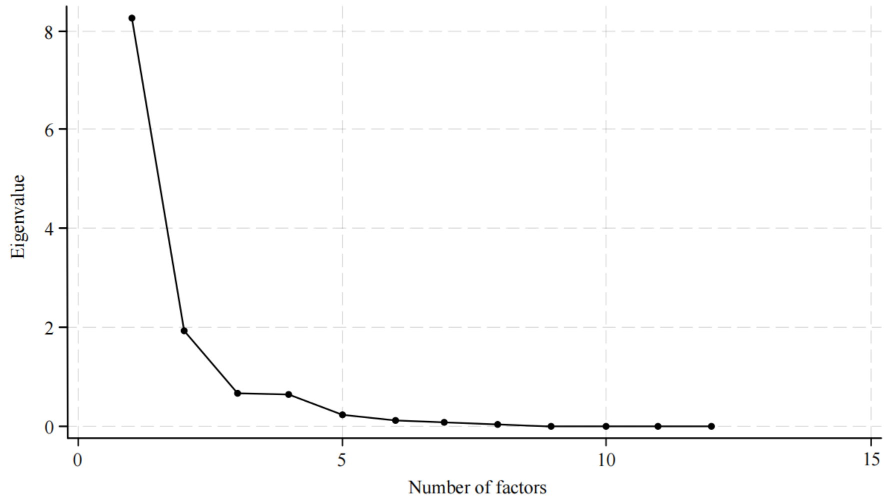

Figure 1. Scree Plot

3.1.3. Rotated Component Matrix and Factor Score Calculation

After extracting the principal components, this study used Stata18, an advanced statistical analysis tool, to perform orthogonal rotation on the extracted components, further deepening the factor analysis process. Through this procedure, the rotated component matrix was obtained, which details the weight distribution and interrelationships of each factor within the rotated principal components. The specific results are shown in Table 4. In this matrix, indicators such as the full-time equivalent of R&D personnel (X4), show significantly high loadings on the first common factor F_1, indicating that they can be summarized as factors representing innovation input and output. In contrast, the number of technology business incubators (X6) and the proportion of enterprises with R&D activities (X7) show relatively high loadings on the second common factor F_2, which can be interpreted as a factor reflecting the innovation environment.

Table 4. Component Matrix after Rotation

Variable	Ingredient	
	1	2
X_1	0.924	-0.011
X_2	0.676	-0.119
X_3	0.723	0.205
X_4	0.993	0.023
X_5	0.984	0.027
X_6	0.095	0.945
X_7	-0.03	0.955
X_8	0.97	0.047
X_9	0.961	-0.115
X_{10}	0.983	0.016
X_{11}	0.933	0.19
X_{12}	0.88	0.149

Table 5. Component Score Coefficients

Variable	Ingredient	
	1	2
X_1	0.113	-0.03
X_2	0.086	-0.08
X_3	0.083	0.088
X_4	0.121	-0.015
X_5	0.12	-0.012
X_6	-0.014	0.491
X_7	-0.029	0.5
X_8	0.117	-0.002
X_9	0.121	-0.086
X_{10}	0.12	-0.018
X_{11}	0.109	0.074
X_{12}	0.104	0.054

3.1.4. Factor Scores and Ranking

The factor score coefficient matrix was obtained using Stata18 software (see Table 5). Based on this matrix, the common factor scores were calculated according to the methods described in Equations (8) and (9), and sort them accordingly.

$$\begin{aligned} F_1 = {} & 0.113X_1 + 0.086X_2 + 0.083X_3 + 0.121X_4 + 0.12X_5 - 0.014X_6 - 0.029X_7 \\ & + 0.117X_8 + 0.121X_9 + 0.12X_{10} + 0.109X_{11} + 0.104X_{12} \end{aligned} \tag{8}$$

$$\begin{aligned} F_2 = {} & -0.03X \ -0.08X_2 + 0.088X_3 - 0.015X_4 - 0.012X_5 + 0.491X_6 + 0.5X_7 \\ & -0.002X_8 - 0.086X_9 - 0.018X_{10} + 0.074X_{11} + 0.54X_{12} \end{aligned} \tag{9}$$

$$F = (68.92F_1 + 16.13F_2)/85.05 \tag{10}$$

According to Equation (10), the variance contribution rate corresponding to each factor score is regarded as a weight factor. The scores of each factor are weighted, and these weighted scores are accumulated to obtain the comprehensive score of each city in Anhui Province. Based on this, they are ranked. The specific results are presented in Table 6.

As shown in Table 6, regarding the F_1 factor ranking, Hefei (3.462), Wuhu (0.974), and Chuzhou (0.176) lead the list, indicating these cities excel in technological innovation investment and output, possessing abundant talent reserves and numerous high-tech industries. For the F_2 factor ranking, Ma'anshan (1.993), Chizhou (1.229), and Chuzhou (0.997) top the list, demonstrating superior technological innovation environments and potential innovation capabilities through their numerous science and technology business incubators and a high proportion of R&D-active enterprises. In the comprehensive score ranking, Hefei leads with a score of 2.769, highlighting its dominant position in Anhui's technological innovation capabilities. Wuhu (0.759) and Ma'anshan (0.337) follow closely behind. The results show that the innovation ability is concentrated in the central cities, while the innovation ability in the peripheral cities is weakened.

Table 6. Technology Innovation Capacity and Ranking of 16 Cities in Anhui Province

City	Total score	Overall ranking	F1 score	F1 ranking	F2 score	F2 ranking
Hefei	2.769	1	3.462	1	-0.194	10
Wuhu	0.759	2	0.974	2	-0.157	9
Ma'anshan	0.337	3	-0.050	4	1.993	1
Chuzhou	0.332	4	0.176	3	0.997	3
Bengbu	0.070	5	-0.073	5	0.681	5
Xuancheng	-0.221	6	-0.267	8	-0.026	8
Fuyang	-0.224	7	-0.278	9	0.008	7
Anqing	-0.247	8	-0.348	10	0.185	6
Chizhou	-0.255	9	-0.602	15	1.229	2
Tongling	-0.285	10	-0.248	6	-0.442	12
Huangshan	-0.327	11	-0.577	14	0.743	4
Lu'an	-0.458	12	-0.443	12	-0.526	13
Suzhou	-0.466	13	-0.444	13	-0.559	14
Huaibei	-0.524	14	-0.254	7	-1.679	15

Bozhou	-0.554	15	-0.607	16	-0.330	11
Huainan	-0.705	16	-0.421	11	-1.922	16

3.2. Cluster Analysis

This paper uses Stata18 software to perform average linkage clustering on the scientific and technological innovation development status of the 16 cities in Anhui Province, resulting in the cluster analysis dendrogram shown in Figure 2.

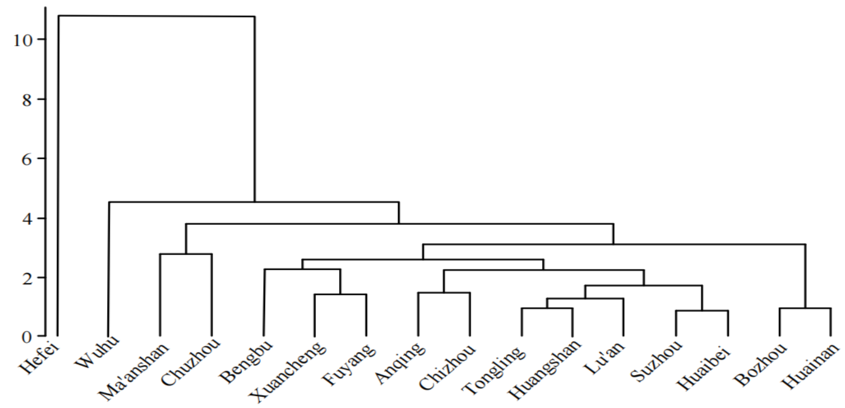

Figure 2. Cluster Analysis Tree Diagram of Scientific and Technological Innovation Capability in 16 Cities of Anhui Province

Figure 2 categorizes 16 cities in Anhui Province into four tiers based on their technological innovation capabilities, with classification details presented in Table 7. Hefei, the provincial capital, ranks first in the first tier, demonstrating the strongest technical innovation capacity. Leveraging its political status and geographical advantages, Hefei has attracted top-tier innovators and cutting-edge tech enterprises, while making substantial investments in R&D. The second tier includes Wuhu City, which closely follows Hefei in innovation metrics. Wuhu's significant funding in scientific research, particularly in R&D activities, highlights its robust innovation capabilities, driving technological advancement and industrial growth. However, the city still has room for improvement in establishing R&D institutions. The third tier comprises Ma'anshan and Chuzhou. While Ma'anshan government provides some support for innovation, its R&D funding falls short of the top-tier cities, requiring increased investment from both authorities and businesses to stimulate innovation. Chuzhou lags behind in innovation ecosystem development, with fewer tech business incubators compared to its peers. The fourth tier includes 12 cities: Bengbu, Chizhou, Huangshan, Fuyang, Anqing, Suzhou, Lu'an, Bozhou, Xuancheng, Tongling, Huaibei, and Huainan. These cities face challenges including insufficient R&D personnel, low R&D expenditure as a percentage of GDP, limited high-tech enterprises, low new product sales revenue, and underwhelming patent applications and academic publications.

Table 7. Clustering analysis results of 16 cities in Anhui Province

Class	Quantity	City
first	1	Hefei
second	1	Wuhu
third	2	Ma'anshan, Chuzhou
fourth	12	Bengbu, Chizhou, Huangshan, Fuyang, Anqing, Suzhou, Lu'an, Bozhou, Xuancheng, Tongling, Huaibei, Huainan

4. Conclusions and Suggestions

4.1. Conclusions

Using factor analysis, we conducted an in-depth evaluation of innovation capabilities in 16 prefecture-level cities of Anhui Province in 2022, examining three dimensions: innovation investment, innovation environment, and innovation output. The analysis identified critical factors reflecting innovation input-output relationships and innovation environment factors. Notably, innovation input-output factors play a significant role in influencing technological innovation capabilities, while innovation environment factors also exert measurable impacts. The ranking of comprehensive factor scores revealed substantial disparities between less economically developed regions (e.g., Huainan, Bozhou, Huaibei) and more developed areas (e.g., Hefei, Wuhu, Ma'anshan). Furthermore, cluster analysis categorized the 16 cities into four distinct groups, highlighting regional variations in technological innovation capabilities and identifying potential areas for improvement. The clustering results indicate that while Anhui has made progress in technological innovation, further efforts are needed to narrow regional disparities.

4.2. Suggestions

4.2.1. Strengthen industrial agglomeration and vigorously cultivate innovation entities

On the path of enhancing regional innovation capabilities, innovation entities are undoubtedly the most valuable resources. Therefore, strengthening and expanding innovation entities, especially innovative enterprises, should be placed at the core of scientific and technological innovation work. Cities in Anhui Province should closely follow the development directions of the province's ten strategic emerging industries, leverage their geographical advantages, effectively undertake industrial transfer, actively introduce upstream and downstream enterprises of leading industries, and pay special attention to and support those enterprises with unicorn potential, with the aim of cultivating more leading enterprises with a scale of hundreds of billions of yuan. This will lead the coordinated development of the entire upstream and downstream of the industry and form an industrial agglomeration effect [10].

4.2.2. Optimize the innovation environment and smooth the chain of innovation cooperation

Governments play a pivotal role in policy-making and social governance, with their responsibility to optimize the innovation ecosystem being particularly crucial. Therefore, cities in Anhui Province should thoroughly optimize the structure and methods of fiscal investment in science and technology, ensuring funds are more effectively allocated to corporate R&D, incubator platform development, venture capital funds, cultivation of technology service intermediaries, and the thriving growth of technology markets. To promote cross-regional collaboration and exchange, the Anhui provincial government should proactively establish close cooperation mechanisms among government, industry, academia, and research institutions, increase financial support for joint R&D projects in the Yangtze River Delta region.Meanwhile, Anhui Province needs to further expand international scientific and technological cooperation channels to attract globally leading technological assets and professionals.

References

[1] Liu Z W, Jiang X R, Zhang X P. Evaluation Index System and Model Construction of Regional Technological Innovation Capability in China [J]. Research on Technology Economics and Management, 2009, (01):32-35.

[2] Zhang X P, Liu Z W, Zhang F. Construction and Evaluation of the Indicator System for Regional Technological Innovation Capability [J]. Research on Science and Technology Management, 2010,30(21):69-74.

[3] Bei S H, Wang Y, Shen J. Evaluation of Jiangsu Province's Technological Innovation Capability Based on Factor Analysis [J]. Science and Technology Management Research, 2021,41(12):77-82.

[4] Zhang J W, Hu Z Y, Ma H D, et al. Coupling and coordination analysis of innovation capacity and economic development in Henan Province [J]. Journal of Northwest Normal University (Natural Science Edition), 2022,58(01):113-119.

[5] Sun L J, Liu X S. Evaluation of Urban Technological Innovation Capacity in China [J]. Journal of Daqing Petroleum Institute, 2007, (02):88-91+130.

[6] Wang J, Wang Y. Comprehensive Evaluation of Innovation Capacity in Anhui's High-tech Industry [J]. China Science and Technology Resource Guide, 2020,52(02):105-110.

[7] Yang S C, Du H Y. Research on the independent innovation capability of Henan Province based on principal component analysis [J]. Nandu Xuetan, 2021,41(04):110-115.

[8] Shen J, Meng Y, Yang B Z. Evaluation of service innovation capability in China's high-tech service industry [J]. Technology and Economics, 2014,33(01):39-47.

[9] Zhou Y, Li H X. Research on the Evaluation of Regional Science and Technology Innovation Competitiveness [J]. Technology and Innovation Management, 2018,39(01):27-33.

[10] Zuo R, Zhao X Y. Evaluation and Analysis of Regional Innovation Capacity in Anhui Province [J]. China Science and Technology Resource Guide, 2023,55(03):77-84+110.