

Study on the Comprehensive Evaluation of Population Quality in Anhui Province Based on the Principal Component Analysis Model

Chenpeng Wu, Ze Wang*, Chun Miao

School of Big Data and Artificial Intelligence, Chizhou University, Chizhou, China

*Corresponding author: 18856611989@163.com

Abstract: Over the past 40 years since the comprehensive implementation of family planning, the rapid population growth has been effectively controlled and the population has been regenerated. The historic change in production type has effectively eased the pressure on resources and the environment, and effectively promoted economic development. This paper uses the population development data of Anhui province in 2020. Through the Anhui Statistical Yearbook, 10 indicators, including birth rate, death rate, illiteracy rate, enrollment rate, dependency rate, unmarried rate, dependency ratio, divorce rate and widowed rate, were collected. After data preprocessing, the data set was systematically clustered by Matlab software, and the quality of Anhui population development was comprehensively evaluated by principal component analysis (PCA). Through the analysis, we find that the top three cities are Bozhou, Fuyang and Suzhou. And provide some reference suggestions for the population development of each city.

Keywords: PCA; population level; comprehensive evaluation

1. Introduction

Population development is the process of the subject of social life, with the progress of social production mode, the change of social and economic conditions, its quantity, quality and structure and its relationship with the outside constantly from lower to higher movement. Population development means the development of the population, especially the ability of the labor force to conquer the nature, the overall improvement of the level of science and technology and culture and education, and the development of the ability to understand and apply the laws of nature and society to constantly transform nature and society. Population development has both quantitative development and change, and quality development and leap [1-2].

According to the statistics of Anhui Population Development Report in 2021, the permanent resident population of Anhui in 2021 was 6,113,10,000, an increase of 100,000 compared with the 20 "seven pu" in 2020, an increase of 0.16%. In 2021, the overall population development of the province will be stable, the population growth rate will slow down, the level of urbanization will continue.

To improve, the quality of the population will be steadily improved, and the degree of aging will be further deepened. Therefore, it is of great practical significance to comprehensively evaluate the population development of all prefecture-level cities for the population development of Anhui Province.

For the comprehensive evaluation of population development quality of prefecture-level cities in Anhui Province, we collected the index data of population development of prefecture-level cities in Anhui Province in 2020 by consulting the 2020 Statistical Yearbook of Anhui Province. In order to more comprehensively reflect the quality of population development in prefecture-level cities, we determine the 10 indicators of population urbanization, birth rate, death rate, illiteracy rate, enrollment rate, dependency rate, the male to female ratio, unmarried rate, divorce rate, and happiness index. The 16 prefecture-level cities were clustered by unsupervised learning. Then the principal component analysis (PCA).

2. Symbol description

Symbol	Explain
x_1	Represents the population urbanization rate
x_2	A birth rate
x_3	Indicates mortality
x_4	Indicates the illiteracy rate
x_5	It means the enrollment rate of compulsory education
x_6	Represents the male to female ratio (%)
x_7	Show support ratio
x_8	The rate of unmarried
x_9	Express divorce rate
x_{10}	Meaning the widowed rate

3. Principal component model establishment

This paper needs to make a quantitative analysis of the population situation of Anhui Province in 2020, and analyze the population quality level of each city in Anhui Province based on the results.

Then analyze the data, because the variables is more, and the correlation between these variables, so the principal component analysis of these high dimensional space variable reduction, and constantly adjust is used to reduce the number of variables and types, finally select cities population urbanization rate, birth rate, mortality, illiteracy rate 10 indicators.

Step 1: Index standardization

In practical application, the magnitude of the indicators is often different, so the original p indicators are standardized first, in order to eliminate the influence on the dimension.Do the following data transformation:

$$\bar{\alpha}_{ij} = \frac{\alpha_{ij} - \mu_j}{s_j} (i, j = 1, 2 \dots, p) \tag{1}$$

$$x_j = \frac{1}{n} \sum_{i=1}^{36} a_{ij}; s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{36} (a_{ij} - \bar{x}_j)^2} (j = 1, 2 \dots, p) \tag{2}$$

Step 2: Calculate the correlation coefficient

The correlation coefficient matrix is calculated based on the standardized data:

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix} \tag{3}$$

$r_{ij}(i,j= 1,2,\dots\text{And } p)$ is the x of the original variable itake part in x_j The correlation coefficient between them is calculated as:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (4)$$

Step 3: Calculate the eigenvalue feature vector

The eigenvalues of the correlation coefficient matrix R and the corresponding eigenvectors, $0 \leq \lambda_p \leq \dots \leq \lambda_2 \leq \lambda_1$ among $\mu_1, \mu_2, \dots, \mu_j$ among $\mu_j = [\mu_{1j}, \mu_{2j}, \dots, \mu_{jj}]^T$

Step 4: Write down the principal component expression

According to the above calculation, the principal component mathematical expression is obtained as follows:

$$\begin{aligned} y_1 &= \mu_{11}x_1 + \mu_{12}x_2 + \dots + \mu_{1j}x_j \\ y_2 &= \mu_{21}x_1 + \mu_{22}x_2 + \dots + \mu_{2j}x_j \\ &\vdots \\ y_j &= \mu_{j1}x_1 + \mu_{j2}x_2 + \dots + \mu_{jj}x_j \end{aligned} \quad (5)$$

Step 5: Calculate the principal component contribution rate and the cumulative contribution rate

Select p principal components to calculate the comprehensive evaluation value.

(1) Calculate the eigenvalues (j = 1, 2, ..., p) Information contribution rate and cumulative contribution

$$b_j = \frac{\lambda_j}{\sum_{k=1}^m \lambda_k} \quad (j = 1, 2, \dots, m)$$

rate.fit

Primary component, yjThe information contribution rate of. At the same time, yes, the main component is

$$\alpha_p = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^m \lambda_k} \quad y_1, y_2, \dots, y_p \quad (6)$$

The cumulative contribution α_p rate. When close to 1 (generally = 0.85, 0.90, 0.95), the first p indicator variables are selected y_1, y_2, \dots, y_p As p principal components, instead of the original m index variables, which enables a comprehensive analysis of p principal components.

$$Z = \sum_{j=1}^p b_j y_j$$

(2) Calculate the comprehensive score:, in the formula: b . For the informative contribution rate of the j-th principal component, according to the corresponding values, the comprehensive score value can be evaluated.

4. The comprehensive evaluation model solution

This paper takes 16 cities as the main research objects, and selects 10 indicators: population urbanization rate, birth rate, death rate, illiterate rate, compulsory education rate, enrollment rate, male and female ratio (%), dependency ratio, unmarried rate, divorce rate, widowhood rate, to understand the impact of different indicators on the population quality level of each city in Anhui Province , as shown

in Figure 1.

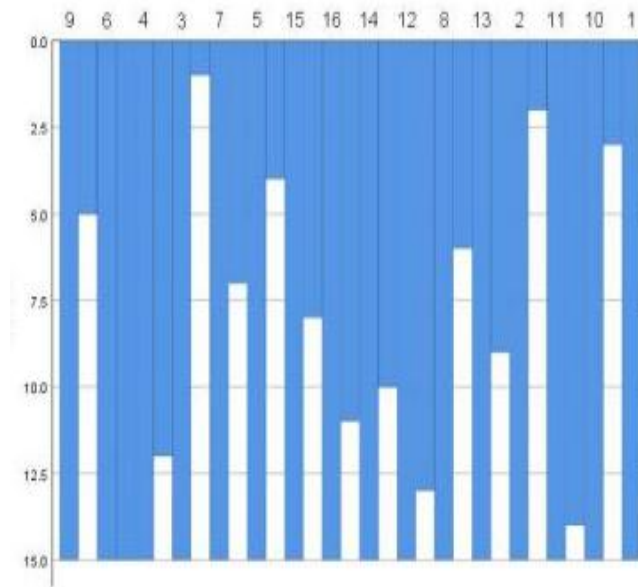


Figure 1: Ice wall chart

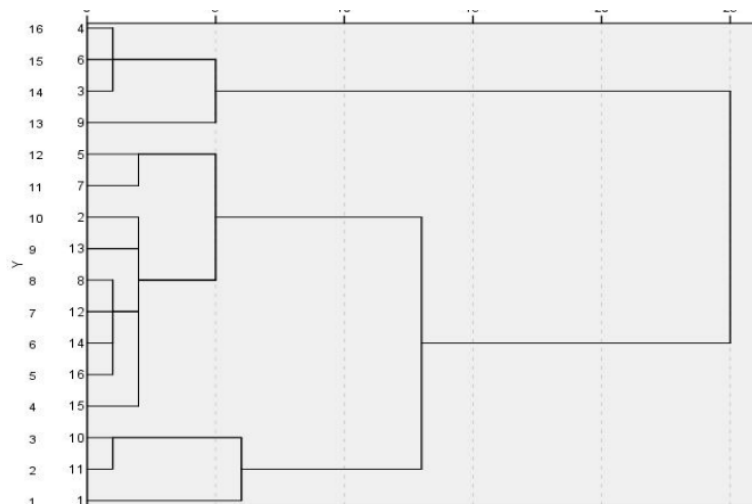


Figure 2: Fig.A dendrogram

Figure 2 shows the classification of the 16 urban areas in Anhui Province, showing the classification status and the distance between the categories[3]. Take Bozhou City, Suzhou City and Fuyang City as an example, they are closest and highly similar, so they are combined into a class and follow this step, and finally all individual provinces and cities gather into a class. At this time, the distance between classes is already very large, and the clustering process is completed.

Comprehensive results, the central city of Anhui province population quality level is higher than the southern cities and northern cities, especially in Hefei, Maanshan and Wuhu three cities, population quality are at a higher level, belong to the prosperity of economic development in Anhui province, followed by the southern region, population quality development level is better, and northern regions due to the influence of geography, policy and other factors, lead to its population quality compared with other cities is still at a lower level.

Table 1: KMO and Bartlett tests

KMO and Bartlett tests		
Number of KMO sampling suitability	Approximate chi square	0.619 102.283
Bartlett sphericity test	free degree conspicuousness	45 0.000

In Table 1 KMO and Bartlett tests, KMO is 0.619, and the observed value of the sphericity test statistics of Bartlett (Bartlett) is 102.283. If the significance level is $\alpha = 0.05$, it is significantly greater than the probability p-value, and the null hypothesis should be rejected. Therefore, there is a significant difference between the correlation coefficient matrix and the unit array, which further verifies that the above data is feasible.

Table 2: Shows the factorial variance

Communality		
	initial	Draw
x_1	1.000	0.835
x_2	1.000	0.711
x_3	1.000	0.751
x_4	1.000	0.834
x_5	1.000	0.794
x_6	1.000	0.703
x_7	1.000	0.924
x_8	1.000	0.842
x_9	1.000	0.758
x_{10}	1.000	0.895

Table 3: Total variance interpretation

Total variance interpretation						
ingredient	Initial eigenvalue		The sum of the load squares was extracted			
	amount to	variance percentage	accumulate%	amount to	variance percentage	accumulate%
1	4.001	40.006	40.006	4.001	40.006	40.006
2	2.791	27.914	67.920	2.791	27.914	67.920
3	1.256	12.555	80.475	1.256	12.555	80.475
4	0.725	7.247	87.722			
5	0.427	4.269	91.991			
6	0.376	3.763	95.754			
7	0.198	1.984	97.738			
8	0.135	1.354	99.091			
9	0.061	0.614	99.705			
10	0.029	0.295	100.000			

Table 2 and Table 3 shows that the cumulative contribution rate of the first three principal components has reached 80.475%, so the first three principal components are extracted for analysis.

Figure 3 gives the eigenvalues under the number of different components, 10 variables in the first fraction load is high, means that they have high correlation with the first principal components, to the second and third fraction has been flat, after the fourth to 10 principal components more stable, further validate the extraction of three principal components is reasonable[4], as shown in Table 4 and Table 5.

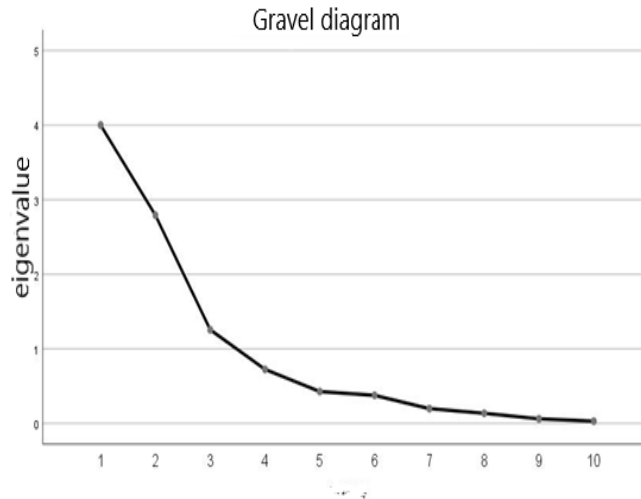


Figure 3: The scree diagram

Table 4: Component matrix A

Component matrix			
	Ingredient		
	1	2	3
x_1	-0.911	-0.052	0.044
x_2	0.643	0.530	0.126
x_3	0.540	-0.586	-0.342
x_4	0.736	0.532	-0.096
x_5	0.086	-0.118	0.879
x_6	0.786	0.229	-0.182
x_7	0.930	0.198	0.139
x_8	-0.426	0.811	0.061
x_9	0.430	-0.561	0.508
x_{10}	0.258	-0.897	-0.156

Table 5: The component score coefficient matrix

	Ingredient		
	1	2	3
x_1	-0.228	-0.019	0.035
x_2	0.161	0.190	0.101
x_3	0.135	-0.210	-0.272
x_4	0.184	0.190	-0.077
x_5	0.021	-0.042	0.700
x_6	0.196	0.082	-0.145
x_7	0.233	0.071	0.111
x_8	-0.107	0.290	0.049
x_9	0.108	-0.201	0.405
x_{10}	0.064	-0.321	-0.124

By calculation, the first, second and third main components of the population development level of each prefecture-level city in Anhui Province are:

$$F_1 = -0.228x_1 + 0.161x_2 + 0.135x_3 + 0.184x_4 + 0.021x_5 + 0.196x_6 + 0.233x_7 - 0.107x_8 + 0.108x_9 + 0.064x_{10}$$

$$F_2 = -0.019x_1 + 0.190x_2 - 0.210x_3 + 0.190x_4 - 0.042x_5 + 0.082x_6 + 0.071x_7 + 0.290x_8 - 0.201x_9 - 0.321x_{10}$$

$$F_3 = 0.035x_1 + 0.101x_2 - 0.272x_3 - 0.077x_4 + 0.700x_5 - 0.145x_6 + 0.111x_7 + 0.049x_8 + 0.405x_9 - 0.124x_{10}$$

Table 6: Comprehensive ranking of population development quality of each prefecture-level city in Anhui Province

City	F 1	F 2	F 3	F overall	ranking
Bozhou city	32.02015	14.24961	62.93168	30.67870739	1
Fuyang city	31.44215	13.31935	62.7907	30.04670056	2
the City of Suzhou	31.27095	12.10043	62.42739	29.48211134	3
Lu'an City	28.62626	11.50256	61.42601	27.80376729	4
Bengbu City	26.91073	11.93559	62.90428	27.33176712	5
Huainan City	25.63062	10.78011	62.15936	26.17838259	6
Anqing City	24.60186	11.22315	62.14336	25.8181411	7
Huaibei City	22.38471	11.02745	63.22923	24.8174704	8
Chuzhou	22 71122	11 02606	61 44055	24 70024991	9
Huangshan city	23.04857	9.01509	62.16429	24.28332991	10
Chizhou city	22.06658	10.51586	61.85857	24.26802935	11
Xuancheng City	22 8707	9 27478	61 03736	24 1091704	12
Tongling city	20.34936	9.86246	61.94444	23.201114	13
Ma'anshan City	19.91386	10.03982	62.32189	23.10502327	14
Wuhu City	18 87615	10 55474	62 14254	22 73978079	15
Hefei City	15.03401	11.89656	62.8204	21.40094814	16

By observing the contribution of each indicator to the principal component, we conclude that:

(1) In the expression of the first principal component, we can see that only the coefficient of the seventh item is relatively large, far exceeding the coefficient of other indicators, so it can be regarded as the influence of the dependency ratio on the quality level of the population, indicating that the dependency ratio also occupies a large proportion in the population quality level, as shown in Table 6.

(2) In the expression of the second principal component, we can see that only the coefficient of the eighth term is relatively large .

The coefficients far exceed the other indicators and thus can be seen separately as an effect on the population quality level, indicating that it also occupies a large proportion of the population quality level.

(3) In the expression of the third principal component, we can see that the coefficients of the fifth and ninth items are larger than the other items, indicating that the compulsory education enrollment rate and divorce rate have a greater impact on the population quality level of each city in Anhui Province. The fifth coefficient is larger, far exceeding the coefficient of other indicators, indicating that the enrollment rate of compulsory education occupies a larger proportion of the population quality level.

5. Results

Suggestions: Effective measures should be taken to solve the problems in the improvement of the population quality level in Anhui province, so as to promote the rapid improvement of the population quality level of all cities in Anhui Province.

(1) Give high priority to the development of education. We must increase investment in education and pay attention to the reform and training systems. We will deepen reform of the education and training systems, train talents at different levels, especially those in short supply, so that the hierarchy of talents meets the needs of economic and scientific and technological development, and transform abundant human resources into advantages in human resources[5]. At the same time, we must also focus on the countryside. Not only that, but also must take strong measures to support, strengthen the development of rural vocational education. Strive for free education in the frontier counties, especially the investment in hardware facilities and equipment, and expand the scale of vocational schools, strengthen the provision of teachers; increase the treatment of teachers in vocational schools, encourage talents to vocational

schools; increase the average public funds of vocational schools, and increase the professional training funds.

(2) The current comprehensive three-child policy implemented by the government will stimulate the pent-up fertility willingness of some urban residents to some extent. This can not only realize the unity of national policies and the will of the masses, but also improve the quality of the population and effectively prevent the phenomenon of "population reverse elimination". The steady implementation of the comprehensive three-child policy can fully predict the role of delaying the speed of the aging population and promoting sustained and stable economic development.

(3) Actively build a social welfare system that is shared by the whole people to benefit the fruits of China's economic and social development

The people (especially the vast number of rural residents), this is a basic project to improve the quality of the population. It is of great significance to actively promote the reform of the household registration system, build the integrated social security system and accelerate the reform of the medical and health old-age insurance system for the citizens to protect the physical quality of citizens and then improve the scientific and cultural quality.

Reflection: In this model, the pre-setting of indicators other than ten indicators, such as population urbanization, birth rate and mortality rate, on the quality evaluation of population development quality in Anhui Province is negligible. And the data with the default statistics covers the whole population of the city, and the impact of the differences in the statistical data methods of each prefecture-level city is not considered. At the same time, the impact of major economic policies, environment, geographical area and other aspects on population development is not considered. Effectively, for the mathematical model of studying the population quality level of each city in Anhui Province, more factors should be considered, and according to the specific research data of different places, such as the population quality level of different age stages and different population quality levels of different genders. Random changing or removing a data, it is not difficult to find that this will have a great change in the results of the population quality level of each city in Anhui Province. In the future, we will continue to study and determine more significant data indicators to design a set of evaluation models that consider more influencing factors to comprehensively evaluate the population quality of each province.

Acknowledgements

Project No.: National College Student Innovation and Entrepreneurship Program under Grant No. 202211306029, National College Student Innovation and Entrepreneurship Program Project under Grant No. 202211306055, the Key Program of the Education Department of Anhui Province under Grant No. KJ2020A0760, the Key Program of Chizhou University under Grant No. CZ2019ZRZ09, the Outstanding Young Talent of Anhui Province under Grant No. gxgnfx2021156.

References

- [1] Zhou Xuexin. *Institutional innovation of comprehensive human development and population development* [J]. *Journal of Guangzhou University (Social Science Edition)*, 2007(02):25-28.
- [2] Yao Tongtong. *Research on influencing factors of Shanghai's economic development* [J]. *China Collective Economy*, 2022(27):34-36.
- [3] Wu Chengxin. *Research on urban residents' income based on principal component analysis and cluster analysis* [J]. *Journal of Huangshan University*, 2021, 23(03):7-10.
- [4] Wu Yanxia. *Research on population quality evaluation and improvement in Xinjiang* [D]. *Shihezi University*, 2014.
- [5] Liu Xinru, Ding Fuxing. *The value and path of improving the quality of the population in the new era* [J]. *Times economy and trade*, 2017(34):78-79.